The interplay of language and visual perception in working memory

Souza, Alessandra S ; Skóra, Zuzanna

Abstract: How do perception and language interact to form the representations that guide our thoughts and actions over the short-term? Here, we provide a first examination of this question by investigating the role of verbal labels in a continuous visual working memory (WM) task. Across four experiments, participants retained in memory the continuous color of a set of dots which were presented sequentially (Experiments 1-3) or simultaneously (Experiment 4). At test, they reproduced the colors of all dots using a color wheel. During stimulus presentation participants were required to either label the colors (color labeling) or to repeat "bababa" aloud (articulatory suppression), hence prompting or preventing verbal labeling, respectively. We tested four competing hypotheses of the labeling effect: (1) labeling generates a verbal representation that overshadows the visual representation; (2) labeling yields a verbal representation in addition to the visual one; (3) the labels function as a retrieval cue, adding distinctiveness to items in memory; and (4) labels activate visual categorical representations in long-term memory. Collectively, our experiments show that labeling does not overshadow the visual input; it augments it. Mixture modeling showed that labeling increased the quantity and quality of information in WM. Our findings are consistent with the hypothesis that labeling activates visual long-term categorical representations which help in reducing the noise in the internal representations of the visual stimuli in WM.

DOI: https://doi.org/10.1016/j.cognition.2017.05.038

Originally published at:
Souza, Alessandra S; Skóra, Zuzanna (2017). The interplay of language and visual perception in working memory. Cognition, 166:277-297.
DOI: https://doi.org/10.1016/j.cognition.2017.05.038

Research Article

The interplay of language and visual perception in working memory

Alessandra S. Souza[a] & Zuzanna Skóra[b]

[a] University of Zurich, Switzerland, [b] Jagiellonian University, Poland

Running-head: Labels in Visual Working Memory

Author Note

Alessandra S. Souza, Department of Psychology, University of Zurich, Switzerland. Zuzanna Skóra, Institute of Psychology, Jagiellonian University, Poland. This research was supported by a grant from the Forschungskredit (FK-13-083) and from the Swiss National Science Foundation (project n° 169302) to Alessandra S. Souza. Correspondence should be addressed to Alessandra S. Souza, Department of Psychology, Cognitive Psychology Unit, University of Zürich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland. E-mail: a.souza@psychologie.uzh.ch

This article has been accepted for publication in Cognition. It has undergone full peer review but has not been through the copyediting, typesetting, pagination, and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as: https://doi.org/10.1016/j.cognition.2017.05.038

Abstract

How do perception and language interact to form the representations that guide our thoughts and actions over the short-term? Here, we provide a first examination of this question by investigating the role of verbal labels in a continuous visual working memory (WM) task. Across four experiments, participants retained in memory the continuous color of a set of dots which were presented sequentially (Experiments 1-3) or simultaneously (Experiment 4). At test, they reproduced the colors of all dots using a color wheel. During stimulus presentation participants were required to either label the colors (color labeling) or to repeat "bababa" aloud (articulatory suppression), hence prompting or preventing verbal labeling, respectively. We tested four competing hypotheses of the labeling effect: (1) labeling generates a verbal representation that overshadows the visual representation; (2) labeling yields a verbal representation in addition to the visual one; (3) the labels function as a retrieval cue, adding distinctiveness to items in memory; and (4) labels activate visual categorical representations in long-term memory. Collectively, our experiments show that labeling does not overshadow the visual input; it augments it. Mixture modeling showed that labeling increased the quantity and quality of information in WM. Our findings are consistent with the hypothesis that labeling activates visual long-term categorical representations which help in reducing the noise in the internal representations of the visual stimuli in WM.

Keywords: categorical memory, continuous features, language, visual working memory;

1. Introduction

The present study is concerned with how visual perception and language interact to form the representations that guide our thoughts and actions over the short-term. The memory system holding information accessible for the moment-to-moment cognition is known as working memory (WM). In WM research, the mainstream strategy has been to study processing of visual and verbal inputs in isolation. In contrast to the laboratory, day-to-day observations suggest a more interactive scenario in which visual inputs and language co-exist and interact. For example, in order to safely change lanes, one has to locate the positions of the other cars, check for traffic signs, and look for potential pedestrians. In each of these steps, one may generate or receive verbal descriptions of the ongoing events. How are these incoming inputs combined in mind to effectively guide action? At the moment, we lack a systematic treatment of the consequences of having both visual and verbal inputs regarding the same event to guide behavior over the short-term. Here, we provide a first examination of this question by investigating the role of verbal labeling for the temporary retention and retrieval of visual inputs varying along a continuous dimension.

The retention of continuous feature values in memory can be studied with the *continuous delayed estimation task* (Prinzmetal, Amiri, Allen, & Edwards, 1998; Wilken & Ma, 2004; Zhang & Luck, 2008). Color reproduction has received the largest attention in the visual WM literature (Allred & Flombaum, 2014). In a typical WM color delayed-estimation task, participants have to retain the precise color-hues of an array of objects. At test, the hue of a target object has to be reproduced using a continuous color wheel. The dependent measure in this task is recall error computed as the distance between the reported value and the target's true value. The more precise the representation of the studied items in memory, the smaller the error in reproducing the target's feature. Furthermore, the distribution of responses in this task can be submitted to mixture modeling to estimate the probability that responses were informed by memory as opposed to guessing, and the variability (imprecision) with which this information was stored (cf. Bays, Catalao, & Husain, 2009; Zhang & Luck, 2008). The sensitivity of this task to the quality of the underlying visual representation makes it a perfect testbed to assess changes in visual WM as a function of verbal labeling.

In standard visual WM tasks, all of the visual stimuli are presented in a one-shot display for a very brief interval (a few hundred milliseconds), and memory is tested shortly after (typically 1 s). The fast pace with which the trial progresses, and the larger number of items displayed simultaneously strongly discourages verbal labeling. This is corroborated by the finding that in change detection tasks (which require the recognition of one of the displayed items) further blocking labeling with the addition of a verbal memory load (cf. Vogel, Woodman, & Luck, 2001) or by asking participants to articulate irrelevant words continuously (aka. *articulatory suppression*) has no impact on performance (Morey & Cowan, 2004, 2005; Sense, Morey, Prince, Heathcote, & Morey, 2016).

To the best of our knowledge, only one study considered how labels affect performance in a continuous delayed estimation task. Donkin, Nosofsky, Gold, and Shiffrin (2015) asked participants to store the precise color of a single dot presented for .1, .5, or 2 s. In color reproduction trials, stimulus offset was followed by a varying retention interval, after which memory was tested with a color wheel. In labeling trials, following stimulus offset, participants were asked to type a label to the color. Three trials later, the label was presented onscreen together with the color wheel, and participants had to pick the color represented by the label. Longer study durations yielded more precise perceptual memory of the stimulus, as well as more precise responding in labeling trials. Modeling of responses in color reproduction trials showed that a mixture of perceptual information, verbal labeling, and of random guessing best fitted the data. Moreover, the model incorporating decay of the visual input over the retention interval, with no decay of the verbal representation, also fitted best. In the study of Donkin et al., the precision of labeling responses was directly assessed in labeling trials, and including this information in modeling improved fitting. Still, this study provided no means to distinguish for the differential impact of labeling because no condition was included that prevented labeling from occurring.

In sum, the extant visual WM literature does not inform us about the possible consequences of allowing people to generate descriptions of the visual events they experience. To tackle this question, we developed an overt labeling protocol to strictly control the labeling opportunities for each item in the memory display. In our experiments, we presented items sequentially, and provided sufficient time after each item to allow for labeling. In the labeling condition, participants were prompted to label each presented item aloud. In the control condition, participants were required to constantly articulate *"bababa"* aloud (articulatory suppression). The articulatory suppression procedure prevents people from articulating and hence generating verbal labels.

Armed with a proper task set-up, we conducted four experiments in which we manipulated further variables to illuminate the space of explanations of the interplay between verbal and visual inputs in memory. In the following, we will delineate four hypotheses of the labeling effect in visual WM that we aimed at distinguishing empirically. We will substantiate the plausibility of these hypotheses with findings from the effects of language on several aspects of cognition, from perception to episodic memory.

1.1. Verbal Recoding

The first possibility is that labeling generates a verbal representation at the expense of the visual one. Storage of the term *"green"* at the expense of the particular greenish hue presented for study should lead to a large loss of precision in recalling this feature from WM. There is evidence that verbalizations can hamper visual long-term memory (LTM). One piece of evidence comes from the verbal overshadowing effect, namely the observation of worse memory for a face (or even color) when in between study and test participants are asked to describe the stimulus (Alogna et al., 2014; Schooler & Engstler-Schooler, 1990). In a related vein, it has been observed that asking people to classify objects as being one out of 2 categories (e.g., lamps vs. chairs) impairs LTM for the studied exemplars, compared to asking for preference ratings (Lupyan, 2008). These studies suggest that verbal descriptions can hamper visual LTM, possibly due to the loss of the visual trace. It is unclear whether a similar effect is observed in visual WM.

1.2. Dual-Trace (Visual + Verbal)

The second possibility is that labeling adds a verbal representation to the visual trace in WM. It follows that participants would have two sources of information: a continuous visual representation, and a verbal label. The joint information from both traces could be combined during recall to yield the final response output; or one of the two representations may dominate depending on the test situation. A dual-trace hypothesis guided the modeling in the study by Donkin et al. (2015): by entering verbal labeling as an additional source of information in mixture modeling, the authors assumed that both visual and verbal inputs co-existed in WM and interacted in guiding reproduction of colors from memory.

Support for this dual-trace hypothesis comes from studies of LTM memory for easy-to-label and hard-to-label drawings performed by Brandimonte and colleagues. They showed that labeling can hamper visual LTM memory, consistent with a verbal overshadowing effect (Brandimonte, Hitch, & Bishop, 1992); however, this effect can be reversed if the visual context for the studied item (e.g. its color) is reinstated at test (Brandimonte, Schooler, & Gabbino, 1997), and the impairing effect depends on the match between the type of verbal description (one label vs. description of features) and the information required at test (global or feature-based) (Brown, Brandimonte, Wickham, Bosco, & Schooler, 2014). They have also shown that verbal descriptions may be beneficial when generated in the presence of the stimulus, but not during a retention interval (Nakabayashi, Mike, Brandimonte, & Lloyd-Jones, 2012). These findings suggest that labeling may yield a verbal representation in memory in addition to the visual input, and that stronger reliance on either type of representation can be varied depending on the retrieval cues presented at test.

1.3. Distinctiveness

The third possibility is that generating a label benefits visual WM because it yields an additional retrieval cue to the labeled item. If participants associate labels with the visual representations, and they remember the pairing of the labels to the spatial locations of items at test, the label can be used to more effectively retrieve the continuous visual representations from WM. If this holds, it would indicate that labeling helps to the extent that it distinguishes between items in memory. According to this account, the LTM impairment observed by Lupyan (2008) could be explained by the lack of distinctiveness of the labels used in this study (only 2 labels for several exemplars of the same category). Richler, Palmeri, and Gauthier (2013) found that LTM for vocally labeled objects (from different categories) was similar to memory for items for which participants made preference ratings, and both conditions yielded better LTM than silent study of the objects. Furthermore, labeling improved rejection of both within-category and between-category lures, hence indicating better memory for the specific exemplars studied. In addition, generating the labels aloud (as opposed to typing them) also played a role, implying a contribution of mode of production to this effect.

1.4. Categorical visual LTM (Visual + Visual)

The last possibility is that labels activate categorical, visual knowledge in LTM. Similarly to the dual-trace hypothesis outlined previously, this hypothesis predicts that labeling allows people to rely on two memory traces. The only difference is that the categorical LTM hypothesis assumes that both traces are visual: one is the visual trace of the studied item; the other is the representation of the visual category in LTM. This hypothesis predicts that labeling helps to activate categorical information, which in turn sharpens the perception and storage of visual information in WM. According to the label-feedback hypothesis (Lupyan, 2012), saying (or hearing) the word *"green"* would have the transient effect of activating visual features related to green and that set it apart from other categories. This would sharpen the perception of the greenish hue presented for study, possibly reducing interference from other colors. This may help to protect representations from forgetting, increase their fidelity, or assist in their short-term consolidation (Ricker, 2015). Although visual categorical knowledge may also be activated by non-verbal means, there is evidence that labels are particularly effective cues to categorical representations (Boutonnet & Lupyan, 2015; Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012; Lupyan & Ward, 2013).

Studies in perception suggest a critical role of labels for the categorical perception of colors. For example, Winawer et al. (2007) asked Russian and English speakers to match one of two shades of blue to a reference blue. In Russian, unlike in English, there is no single term to refer to blue; darker shades are termed *"siniy"* and lighter shades *"goluboy"*. Hence for the English speakers, this task required a within-category discrimination; whereas for Russians, the task was a mixture of within-category and between-category discriminations. Russians, but not English speakers, responded faster for between-category than within-category discriminations, showing evidence of categorical responding. Moreover, this facilitation disappeared under verbal (but not spatial) memory load (see also Athanasopoulos, Damjanovic, Krajciova, & Sasaki, 2011; Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009). The latter finding may suggest that the representation of the item was verbal in nature, hence being more in line with the dual-trace hypothesis described previously. The dual-trace hypothesis predicts impairments by a verbal memory load at any time during the task (from encoding to retention) because the verbal representation needs to be generated and maintained in WM. In contrast, the categorical LTM hypothesis assumes that labels are only needed for activating the categorical code but not afterwards, and a verbal load added during the retention interval should not impair performance. Given that the verbal load in the experiment of Winawer et al. occurred prior to the onset of the color matching task, it discouraged labeling altogether. Hence the absence of categorical bias under verbal load does not distinguish between the source of categorical bias (verbal WM or categorical LTM). Pilling, Wiggett, Özgen, and Davies (2003) showed that a memory advantage for colors spanning two categories disappears under a verbal interference condition (see also Roberson & Davidoff, 2000) only when this test condition discouraged labeling from occurring during encoding. When participants had an incentive to label items in all test conditions, the categorical advantage remained even under verbal interference. This finding is as predicted by the categorical LTM hypothesis.

Some recent reports indicated that visual perception and visual WM for continuous color hues show evidence of categorical bias (Bae, Olkkonen, Allred, & Flombaum, 2015; Bae, Olkkonen, Allred, Wilson, & Flombaum, 2014; Hardman, Vergauwe, & Ricker, 2017). Bae et al. (2015) observed that participants select more often colors close to category centers (as measured in a labeling task) and less often colors at category boundaries both in perceptual and WM tasks. Likewise, Hardman et al. (2017) observed clusters of responses around certain color values which were consistent with participants responding categorically. Both Bae et al. (2015) and Hardman et al. (2017) have incorporated categorical influences (probability that participants make categorical responses) to mixture modeling of the color delayed estimation task. Those models were better able to predict responses in those tasks than models not including categorical influences. Given that in both studies verbal labeling was not controlled for, it is unclear whether there is a contribution of verbal labels to the categorical effects observed. Notwithstanding, those extended mixture models can provide substantial leeway in understanding how verbal labeling affects visual WM, because they allow to quantify how much of the information in WM is continuous as opposed to categorical.

1.5. The Present Study

Our primary goal was to establish the conditions in which labeling is helpful, inconsequential, or harmful to performance in a continuous visual WM task. For that, we will take advantage of the categorical-continuous mixture model developed by Hardman et al. (2017). This model estimates three key parameters: (1) the probability that responses were informed by memory (P^M); (2) the probability that information in memory is continuous (P^O) as opposed to categorical ($1 – P^O$); and (3) the precision of the continuous information in memory (σ^O). Crucially, we will use P^M and P^O to compute a measure of capacity known as K (e.g., Cowan, 2001). K is assumed to reflect the number of items accessible in WM at the time of test. To compute the total number of items in memory (total K), one needs to multiply the probability that information was in memory (i.e., P^M) by the number of studied items (set-size). To separately estimate the number of continuous versus categorical representations in memory, we need to include the estimate of the probability of a continuous representation (i.e., P^O) in the equation (continuous K = $P^M \times P^O \times$ Set-Size; categorical K = $P^M \times [1- P^O] \times$ Set-Size). Estimating continuous K and categorical K is critical for distinguishing between the hypotheses of the labeling benefit.

The verbal recoding hypothesis predicts that verbal labeling does not change the number of items stored in WM (total K), but reduces the number of items stored continuously (continuous K). In contrast, the remaining hypotheses predict that labeling increases total K. However, those hypotheses differ on their assumptions about the reason for gaining this additional information. According to the dual-trace hypothesis, the additional information gained from labeling is categorical and verbal in nature, with no change in the number of continuous representations in WM. Accordingly, this leads to the expectation of an increase in categorical K with no change in continuous K. The distinctiveness hypothesis assumes that labeling helps in the retrieval process. Given recent reports showing storage of both continuous and categorical representations in visual WM (Bae et al., 2015, 2014; Hardman et al., 2017), assistance with retrieval should increase the accessibility of both continuous and categorical information. Lastly, the categorical LTM hypothesis assumes that the activation of categorical representations sharpens visual perception thereby reducing irrelevant color interference. It may also provide additional information. Accordingly, participants may gain both in terms of the continuous representations they store (in their quantity or quality), and in the number of categorical representations.

Across four experiments, we manipulated variables that allowed us to test these predictions. In Experiment 1, we explored the joint effects of labeling (Color Labeling vs. Suppression) and of memory set-size. This experiment allowed us to test whether the labeling effect depends on the load on visual WM. In the subsequent experiments, participants always encoded 4 items to keep a high demand on visual WM. In Experiment 2, we included two additional conditions besides Color Labeling and Suppression: (a) a Position Labeling condition allowing us to assess the effect of a label that helps discriminating between items in the memory array but has no categorical color information; and (b) a Preference Rating condition assessing the contribution of increased attentional demands imposed by labeling. Assessing the impact of a second type of label is essential to distinguish between the distinctiveness and the categorical LTM hypotheses: any label that discriminates between items in the array should yield a benefit according to the distinctiveness hypothesis; the categorical LTM hypothesis, in contrast, predicts a benefit only for labels that carry categorical information.

In Experiment 3, we compared performance yielded by the presentation of visual stimuli (with and without labels) with the performance afforded by memory of only the verbal labels. This manipulation allowed us to test whether performance in the Color Labeling condition could be fully explained by the simple retention of labels in verbal WM (dual-trace hypothesis), or whether labeling helps because it activates representations in visual LTM.

Lastly, in Experiment 4, we assessed the effects of labeling in a more traditional visual WM set-up: instead of the sequential presentation of items used in the previous experiments, all items were presented simultaneously onscreen. To further test the impact of labeling, we varied the retention interval (and hence the time available for labeling), and whether overt labeling was encouraged via instruction. This experiment provided a bridge between results from traditional WM experiments and the method used in our previous experiments (with sequential presentation).

The data and analyses scripts for all experiments reported here are available at the Open Science Framework at: https://osf.io/tf93q/

Across all experiments, we observed that color memory was categorically biased irrespectively of labeling (i.e., even under suppression, and with one-shot brief displays); but labeling increased reliance on categorical information. Mixture modeling showed that labeling improved both the quantity as well as the quality of visual information in WM. More specifically, labeling increased not only categorical K, but also estimates of continuous K and/or continuous precision. This is in line with the predictions of the categorical LTM hypothesis. In sum, verbally activating categorical representations in visual LTM allows people to alleviate the severe constraints imposed by visual WM to the retention of information over the short-term.

2. Experiment 1

In Experiment 1, we set out to investigate the role of labeling while varying the memory load on visual WM. Participants were asked to maintain a sequence of 1, 2, or 4 colored dots, and to reproduce their precise color in a continuous color wheel. This task was performed under two conditions varying in the opportunity to verbally label the colors. In the *Suppression condition*, participants were required to repeat "*bababa*" aloud. In the *Color Labeling condition*, participants were asked to name the colors as soon as they were presented onscreen. The Suppression condition offers a measure of visual WM in the absence of verbal labels. Contrasting this condition against the Color Labeling condition allowed us to assess how labels affect the retention of visual information in WM.

2.1. Method

2.1.1. Participants

For all experiments reported here, participants read and signed an informed consent form prior to the experiment and were debriefed at the end. The experimental protocol is in accordance with the Institutional Review Board of the University of Zurich.

In total, we collected data of 21 students from the University of Zurich (17 women, 4 men; M = 24.4 years old). Participation was compensated with course credit or money (15 CHF per hour). E1 comprised two experimental sessions, and the data was collected in two waves. The first wave (collected in August, 2014) comprised the data of 11 students. For this sample, there were 480 trials per session. One participant did not show up for the second session and was therefore excluded from the analysis. The second wave (collected in October, 2016) comprised 10 students. For this sample, there were 360 trials per session.[1]

2.1.2. Materials and procedure

All experimental tasks were programmed in MATLAB using the Psychophysics Toolbox extension (Brainard, 1997; Pelli, 1997). In Experiment 1, participants were tested in individual booths where they sat approximately 50 cm from the computer screen (viewing distance was unconstrained). Participants wore headphones and were previously informed that their speech would be recorded in order to check for compliance with the experimental instructions.

Figure 1a illustrates the flow of events during the study phase. Each trial was self-started by pressing the spacebar. Thereafter 1, 2, or 4 placeholders (dark grey disks) were shown against a grey background for 500 ms. The disks (radius = 1.6° visual angle) appeared evenly spaced around an imaginary circle (radius = 6.65°) centered in the middle of the screen. The disks were presented at the 90°, 180°, 270° or 360° angle in the imaginary circle. In set-size 1 trials, any of these four locations were equally likely to be used. In set-size 2 trials, the first stimulus location was randomly selected, and the second stimulus location was set to be the one further 90° clockwise in the circle (e.g., 180° and 270°; 360° and 90°). In set-size 4 trials, all four locations were used. As for set-size 2, the first position was randomly selected, and presentation of the subsequent items followed in a clockwise fashion.

[1] We had to reduce the number of trials because participants were taking substantially longer to complete the sessions. Given that the start of the trials was self-paced, individual differences may substantially contribute to the overall duration of the session.

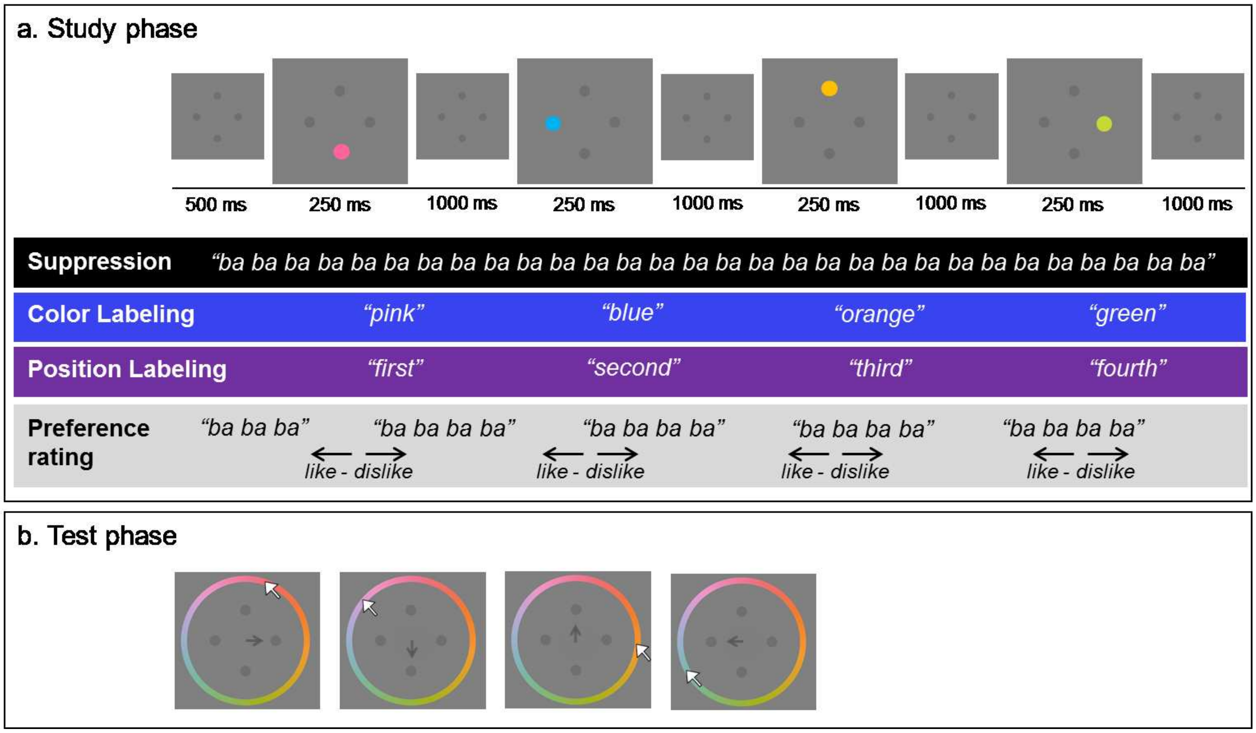

Figure 1. Flow of events in the study (Panel a) and test (Panel b) phases of the continuous color reproduction task used in Experiments 1, 2, and 3. A set-size 4 trial is depicted. During the study phase, different experimental conditions were implemented which varied in the opportunity to use verbal codes and to attend to the items as illustrated by the colored rows in Panel a. Experiments 1 and 3 included the Suppression and Color Labeling conditions, whereas Experiment 2 included all depicted conditions.

Items were studied sequentially by filling one of the placeholders with color for 250 ms, followed by 1000 ms interval in which only the placeholders were onscreen (hereafter the inter-stimulus interval, ISI). The color of each disk was chosen randomly on every trial by selecting among 360 color values that were evenly distributed in a circle in the CIELAB color space. This circle was centered in the color space with L = 70, a = 20, b = 38, and radius = 60 (Zhang & Luck, 2008).[2]

[2] We have not checked that the specified RGB colors were properly rendered by the computer monitors in all experiments reported here. Differences between the intended and rendered colors are however constant across

In two separate sessions (order counterbalanced across participants), participants were required to either repeat *"bababa"* continuously aloud (Suppression condition) or to label the colors (Color Labeling condition). In the Suppression condition, a message was displayed before each trial prompting participants to start with suppression and to press the spacebar to initiate the trial. Participants were allowed to stop articulation in the test phase. This was done to equate articulatory demands between conditions. In the Color Labeling condition, participants were instructed to name the colors as soon as they appeared onscreen. Participants were instructed to use whatever label they found most suitable.

Next, the test phase started with the presentation of a color wheel (randomly rotated from trial-to-trial) and of the mouse-cursor in the center of the screen. A dark-grey arrow (recall-cue) randomly pointed to one placeholder indicating that the color presented at this location should be reproduced on the wheel (henceforth the recall target; see Figure 1b). The arrow remained onscreen until the response was made. Participants indicated the color of the target by clicking with the left-mouse button on a point on the wheel. After responding, the next item was cued to be recalled, and this procedure was repeated until all items were tested. Recall order was randomly determined in each trial. After the recall phase, a 1000-ms blank interval followed before the start of the next trial.

On both experimental sessions participants wore headphones, and their speech was recorded. Set-size was varied randomly between trials, with the constraint that each set-size occurred in an equal number of trials. In the beginning of each session, participants completed

experimental conditions and hence cannot explain performance differences between conditions. We used a custom Matlab code for specifying the colors in CIELAB model, which can be found in the OSF.

9 practice trials which were excluded from subsequent analysis. Participants were instructed to respond as accurately as possible.

2.1.3. Data Analysis

Hardman et al. (2017) created a categorical-continuous mixture model that allows the estimation of the proportion of colors remembered categorically versus continuously. In this model, each category has a specific value (category mean). Because they had no prior information about their participants' categories, their model freely estimated the number of categories and their means for each individual. Their model assumes that representations in memory are either categorical (a few canonical values) or continuous (fine-grained detail about the studied hue). Responses informed by categorical representations cluster around a few canonical values as illustrated in Figure 2a. Responses informed by continuous representations vary linearly with the studied color (see Figure 2b). The storage of the continuous information can be more or less fine-grained, and this is captured by the model parameter called continuous imprecision[3], which in Figure 2b is reflected by the width of the diagonal line. Responses not informed by memory constitute guessing. Guessing can be categorical (random selection of color categories; see Figure 2c) or uniformly distributed on the wheel (continuous guessing; Figure 2d). Response distributions in the task reflect a mixture of these four states, which the mixture model aims at disentangling. [4]

Here we took advantage of this model to assess the impact of verbal labeling in visual WM. Three parameters of the model were of interest: (a) the probability of storage in WM (P^M);

[3] This is the sigma parameter of the von Mises distribution, which has the same meaning as the imprecision parameter in the traditional mixture model proposed by Zhang & Luck (2008).
[4] Scatterplots with experimental data are shown in the Online Supplementary Materials.

(b) the probability that the representation in memory was continuous (P^O) as opposed to categorical (1 - P^O); and (c) the imprecision of the continuous representation in memory (σ^O). Those parameters were allowed to vary across conditions. Other parameters were fixed across conditions: the probability of categorical guessing (P^{AG}), how colors were assigned to categories (category selectivity, σ^S), and the imprecision on the selection of the category (σ^A). The latter captures the fact that categorical responses may deviate slightly from the category center due to motor noise (see width of the categorical bands in Figure 2a). Lastly, given that our labeling data provided information about the color categories used by participants, we fixed the number of categories and their locations for all participants (for category means, see results section on verbal labeling).

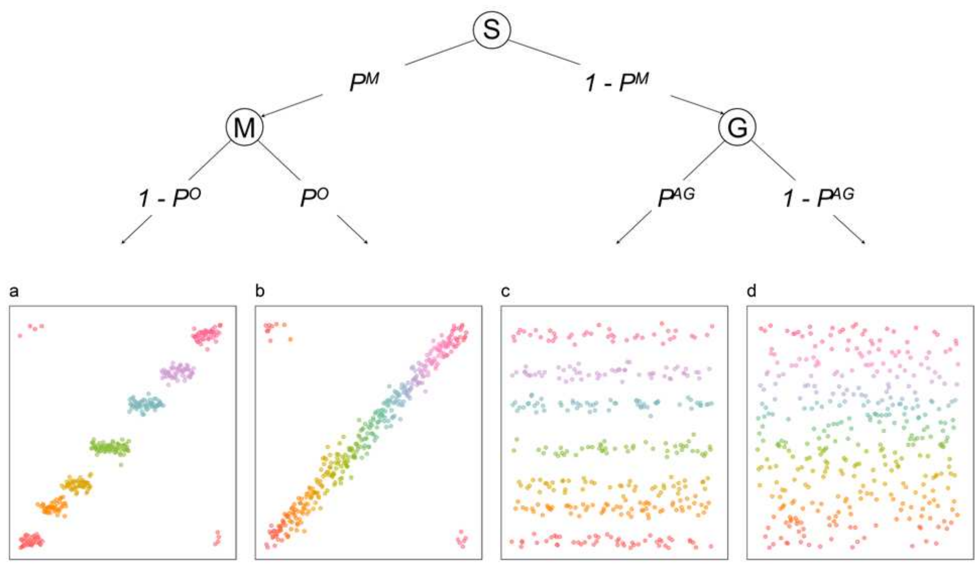

Figure 2. Multinomial process tree for the categorical-continuous model of Hardman et al. (2017). For all scatterplots, the x-axis represents the studied color-hue and the y-axis the response hue. Panel a. Categorical memory: for a range of studied hues, the same categorical response is provided. The width of the categorical bands reflects categorical imprecision. Panel b. Continuous memory: responses vary linearly with the studied hue. The width of the diagonal line indicates the continuous imprecision. Panel c. Categorical guessing: guessing is distributed over categories. Panel d. Random guessing.

Hardman et al. (2017) presented two variants of their model. The between-item model assumes that each response is based on either a categorical or a continuous representation. The within-item model assumes that responses are based on a combination of a continuous and a categorical representation of each item. The latter model is similar in assumptions to the one implemented by Bae et al. (2015). Both models were implemented in a Bayesian Hierarchical Framework, and all parameter values were determined through Bayesian Markov chain Monte Carlo (MCMC) sampling techniques. Hierarchical models assume that the parameter values of individual participants in a given condition are drawn from a population-level normal distribution. Hence in this type of model, we are interested in the estimate of the population-level parameters in each condition, and in assessing whether condition estimates differ from each other. All inferences reported here were based on Bayesian hypothesis testing. Bayesian inference combines prior knowledge about the parameter space (hereafter the “prior”) with the knowledge about the parameter space after seeing the data (the “posterior”). Unlike traditional hypothesis testing, Bayesian inference is not based only on the mean parameter estimate, but also deals with its uncertainty (Kruschke, 2011). One can assess parameter uncertainty by describing the interval covering 95% of the posterior distribution (i.e., its 95% credible interval) alongside its mean value. To compare estimates across conditions, we used Bayes Factors (BF). The BF is the ratio of the likelihood of the hypotheses under comparison (e.g., the Alternative hypothesis over the Null hypothesis; represented by BF_{10}). The BF for the comparison between conditions/factors was performed using the Savage-Dickey density ratio. This method compares the density at the same point (e.g., at 0) in the posterior and in the prior distribution. For details regarding the implementation of the model variants, priors, and BF estimation please refer to the paper by Hardman et al. (2017). For more details regarding the computation of the BFs for factorial designs please refer to Ricker and Hardman (submitted).

2.2. Results

2.2.1. Verbal Labeling Data

For each trial of the Color Labeling condition, we recorded and coded the labels used. Overall, over 60 different labels were used; however most of the labels referred to a set of 7 common color terms (red, orange, yellow, green, blue, purple, and pink) which were also frequently used across all experiments reported here. Figure 3a shows the proportion of verbal responses in which these common terms were used as opposed to other terms (e.g., olive, kiwi, gold, light blue), and to trials in which no labels were spoken or the recorded sound file did not allow for classification (unintelligible category). In the next step, we plotted the proportion of times a given color on the wheel was labeled with one of the common terms (see Figure 3b). As those distributions were bell-shaped, we fitted to them a normal distribution for circular space (i.e., a von Mises). This allowed us to estimate the color-label mean (reflecting the category center) and standard deviation (sigma parameter; see Figure 3c).

As shown in Figure 3c, the mean and standard deviation of the color categories in Experiment 1 were similar across all set-sizes. This indicates that participants did not change how they labeled the colors as a function of memory load. Hence we collapsed the category means across all set-size levels for inclusion in the mixture modeling analysis.

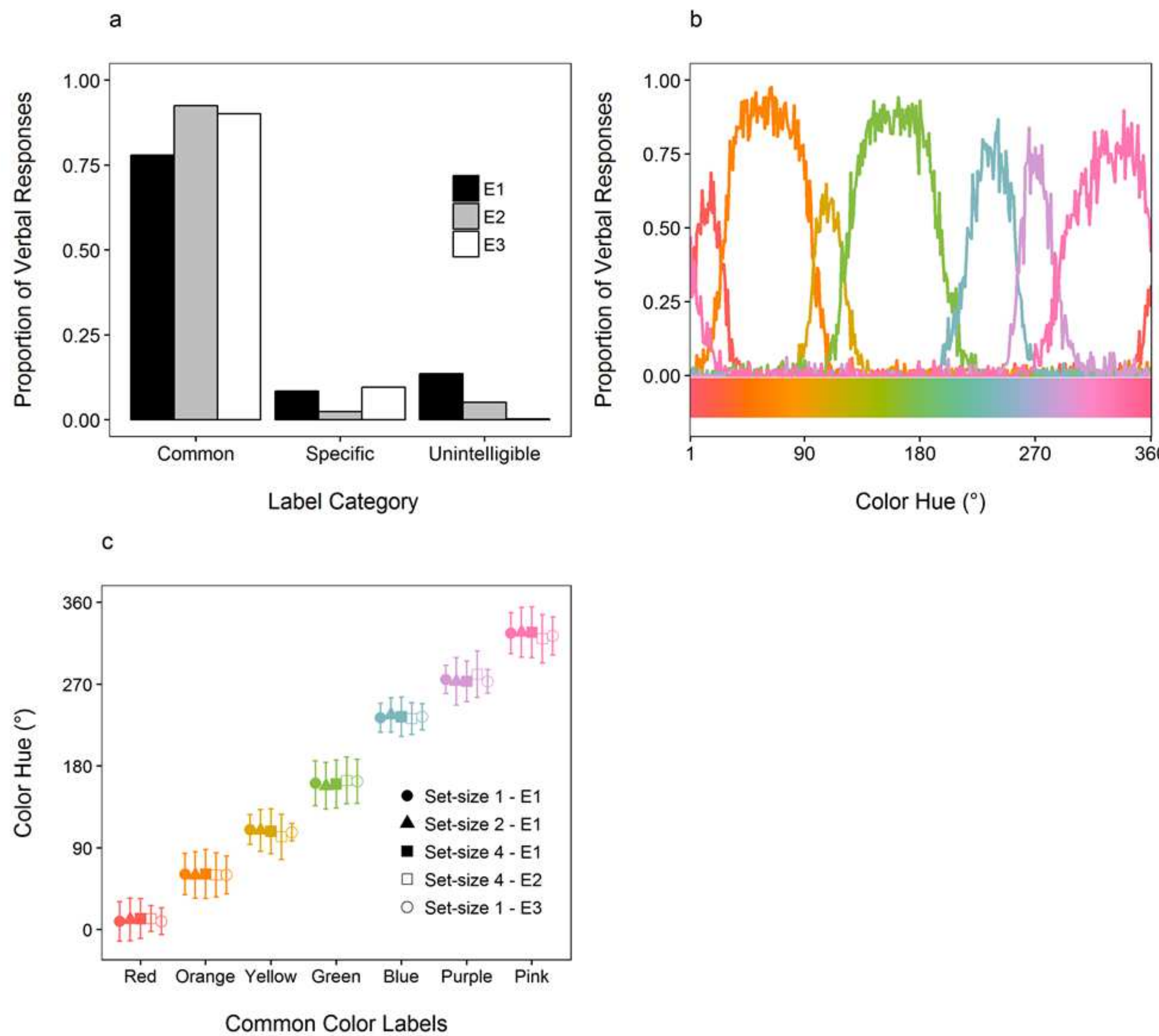

Figure 3. Panel a. Proportion of verbal responses referring to common labels (e.g., red), specific labels (e.g., kiwi), or which were unintelligible across Exps. 1-3. Panel b. Distribution of the labeling responses (for the 7 common terms) over the color-hue space in Exp. 1. Panel c. Mean and standard deviation (sigma) of the von Mises distribution fitted to common label distributions in Exps. 1, 2, and 3.

2.2.1. Mixture Modeling

Model Fitting. We fitted both categorical-continuous model variants to our data using the CatContModel package (Hardman, 2016) implemented in R (R core team, 2014). We fitted the data of all conditions and participants simultaneously using a factorial design with the factors set-size and verbalization. For each model, we ran three parallel chains of 10’000 iterations (with a burn-in of 2’000 iterations). To assess model fit, we used Watanabe-Akaike Information Criterion (WAIC). WAIC is based on the model posterior predictive accuracy and includes a penalty for the effective number of model parameters. Smaller WAIC indicates better model fit. This statistic is similar to other common model fit statistics (AIC, BIC), but it has been considered more appropriate for hierarchical Bayesian models (Gelman, Hwang, & Vehtari, 2014) and it is recommended by Hardman (2016) as more appropriate for comparing CatContModel variants. The between-item model had a smaller WAIC than the within-item model (Δ = -512.2), indicating that this model had a better fitting. This replicates the findings by Hardman et al. (2017). Therefore we will only consider the results of this model here.

We further assessed whether fixing model parameters across verbalization conditions affected the between-item model ability to account for the data. We created two additional models in which we either fixed continuous imprecision (σ^O) or the probability of storage of a continuous representation in memory (P^O) to be the same across the Suppression and Color Labeling conditions (while still allowing for an effect of set-size on those parameters). Fixing either σ^O (ΔWAIC = +578.5) or P^O (ΔWAIC = +334.6) yielded a worse fit compared to the full model. Moreover, to assure that our results do not depend on fixing the category means to particular values, we also fitted a variant in which we allowed the model to freely estimate the categories for each participant. The results of this model can be found on the Online Supplementary Materials. In sum, allowing the model to freely estimate the categories did not substantially change the main pattern of results (if anything the labeling benefit become more substantial). Here we report the estimates of the model in which category means were constrained for all participants.

Finally, to assess the degree that the model captured the data, we performed a posterior predictive check by simulating data based on the model parameters (henceforth predictions). We then computed the absolute distance on the color wheel between the response and true color of the studied item (i.e., measure of recall error) in the observed data and in the predictions (see Figure 4a). For both the empirical data and the predictions, the color labeling condition yielded a smaller recall error, and this labeling benefit grew larger, the higher the load on visual WM. A more detailed predictive check can be found in the Online Supplementary Materials.

Parameter Estimates. The population-level parameter estimates that were fixed across conditions are listed in Table 1. Figures 4b-4d present probability than an item was in memory (panel b), probability that it was stored continuously (panel c), and the imprecision of the continuous representations (panel d), respectively, as a function of set-size and verbalization condition. Table 2 presents the BF_{10} for the main effects of set-size and verbalization, and their interaction. BFs provide a continuous index of the strength of evidence for Alternative hypothesis over the Null. BF_{10} below 1 shows evidence for the Null hypothesis, whereas BF_{10} above 1 shows evidence for the Alternative hypothesis. BFs should be used to update our prior beliefs on the models under consideration. For instance, a BF_{10} = 10 indicates that the Alternative hypothesis is 10 times more likely than the Null, and we should update our belief on this hypothesis over the Null by a factor of 10:1. It is common to consider BF_{10} in the range of 0.3 to 3 as providing ambiguous support for either hypothesis (Kass & Raftery, 1995).

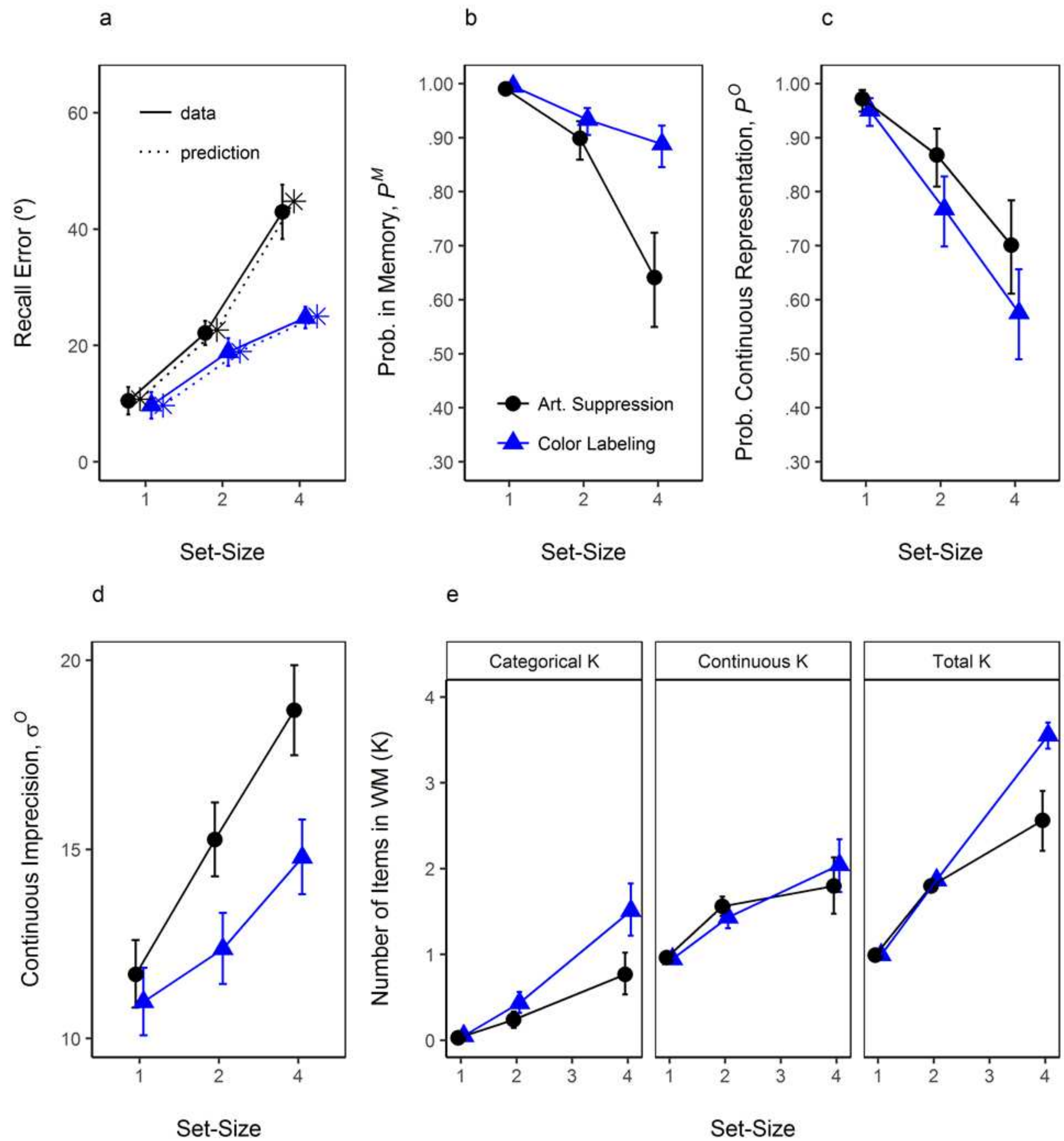

Figure 4. Panel a. Mean distance on the color wheel between study and recalled color in the observed data and in predictions of the mixture model. Error-bars depict 95% within-subject confidence intervals. Panel b. Probability of having the tested item in memory. Panel c. Probability that the memory representation is continuous. Panel d. Imprecision of the continuous memory representation. Panel e. Estimates of the number of items in working memory based on the total information available in memory, the subset of this information that is continuous, and the subset that is categorical. Error-bars in panels b-e depict 95% credible intervals of the parameters.

Table 1

Mean Estimates and 95% Credible Intervals for Parameters of the Mixture Model With No Condition Effect in All Experiments.

Experiment	**Categorical**		
	Guessing (P^{AG})	**Selectivity (σ^{S})**	**Imprecision (σ^{A})**
1	0.01 [0.00, 0.05]	18.8 [17.5, 20.0]	15.2 [13.2, 17.2]
2	0.12 [0.00, 0.67]	22.1 [21.1, 23.2]	23.9 [21.2, 26.7]
3	0.06 [0.00, 0.15]	19.1 [18.1, 20.2]	13.3 [12.3, 14.2]
4 – first response	0.02 [0.00, 0.08]	24.8 [22.0, 28.1]	18.0 [15.9, 20.5]
4 – all responses	0.03 [0.00, 0.08]	23.6 [18.5, 28.9]	13.7 [12.2, 15.3]

Table 2

BF_{10} for the Effects of the Factors Manipulated in Experiment 1.

Term in the Model	**Parameter**		
	Probability Memory (P^{M})	**Probability Continuous (P^{O})**	**Continuous Imprecision (σ^{O})**
Set-Size	6.33×10^{42}	3.48×10^{16}	8.20×10^{22}
Verbalization	1.78×10^{6}	37.10	3.24×10^{31}
Set-Size x Verbalization	1.22×10^{5}	0.003	1070

As shown in Figures 4b-d, increasing set-size decreased the probability that an item was in memory and the probability that it was stored as a continuous representation (i.e., P^{M} and P^{O}), while continuous imprecision (σ^{O}) increased. The set-size effect on all parameters was supported by overwhelming evidence (see Table 2). Labeling the colors increased the probability of having the test item in memory, but decreased the probability that this information was continuous compared to the Suppression condition. Although labeling reduced the probability of storing a continuous representation, it positively affected its quality, as reflected by the reduction in σ^{O}. The main effect of labeling was supported by substantial to very strong evidence (see Table 2). Furthermore, there was evidence for an interaction between set-size and verbalization in the overall probability of storage in memory and in continuous imprecision. This shows that labeling ameliorated the impairing effects of memory load on the total amount of information in memory and on its fidelity.

To tackle the question of how labeling changes the representations stored in WM, we used the estimates of the probability of having an item in memory and the probability that this information was continuous to compute K (total K, continuous K, and categorical K; see Figure 4e). Furthermore, we assessed the verbalization effect (difference between Color Labeling and Suppression conditions) at each set-size level by computing the difference between the posterior estimates of continuous K and categorical K (Figure 5). In combination, Figures 4e and 5 show that labeling yielded a modest reduction in the number of continuous representations in WM when memory load was low (set-sizes 1 and 2). When memory load was high (set-size 4), both continuous K and categorical K increased in the Color Labeling condition compared to the Suppression one.

2.3. Discussion

Experiment 1 showed that labeling continuous colors aided their retention in visual WM compared to Suppression. Our modeling showed that this effect was mainly due to an increase in the number of categorical representations in WM. Two findings indicate that labeling also had an impact on the continuous information stored in WM. First, labeling protected the fidelity of the continuous representations against the impairing effect of memory load. Second, labeling yielded a modest reduction on the number of continuous representations in WM when set-size was low, mainly due to some non-zero probability that participants responded based on the category instead of the continuous information. However, when memory load was high, labeling increased the storage of categorical as well as continuous information in WM. Hence altogether, our data indicates that the storage of continuous information also benefits from labeling.

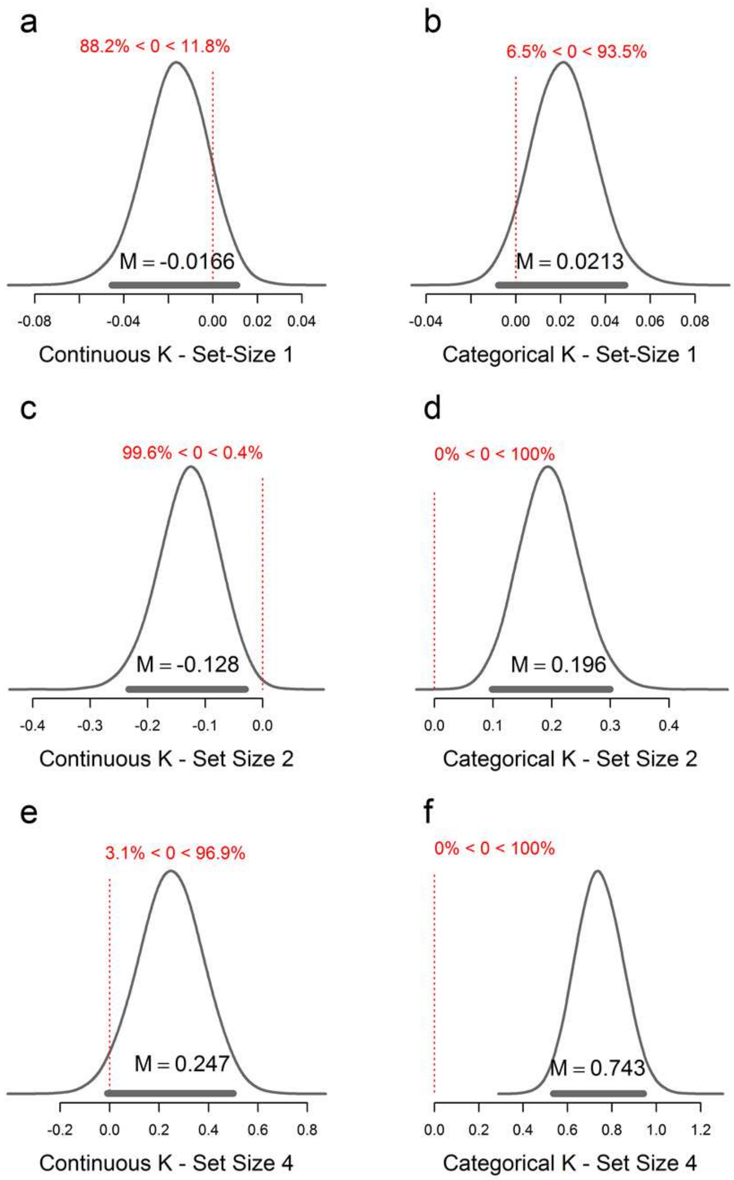

Figure 5. Posterior of the verbalization effect (Color Labeling - Suppression condition) on estimates of continuous K and categorical K at each set-size level in Experiment 1. Each panel presents the percentage of the curve that is above and below 0 (null effect), the mean, and the 95% credible interval of the mean (bar underneath each curve).

Overall, the findings of Experiment 1 are inconsistent with the verbal recoding hypothesis. Total K increased with labeling, contrary to its predictions. Although continuous K tended to decrease slightly under low memory load, this effect was only credible in the set-size 2 condition; and when load was high, continuous K increased. Furthermore, estimates of imprecision decreased indicating higher fidelity of the continuous information stored. All in all, the pattern of findings emerging across cannot be explained by this hypothesis, as labeling helped more than hampered the storage of continuous representations in WM.

3. Experiment 2

Experiment 1 provided first evidence that labeling colors aids their retention in WM due (in part) to the availability of more categorical representations (K categorical). Notwithstanding labeling also helped in-the-moment color memory by reducing memory imprecision, and sometimes by increasing the number of continuous representations in WM. This finding leaves open the possibility that labeling helps for reasons beyond color categorization.

We envision two further processes by which labeling could improve visual WM. First, color labeling may help discriminating between items in memory by increasing the number of possible retrieval cues associated with them (H3). Hence it is unclear whether the benefits of assigning a label are constrained to color labeling or could also (partially) arise when providing any other type of label that may help separating items from each other. This alternative account is plausible given recent reports that increasing the number of retrieval cues helps recall. For example, Bae and Flombaum (2013) asked participants to encode either the luminance or the size of a set of objects. When items in the memory display were different not only on the levels of the relevant feature but also on another irrelevant feature (such as color or shape) recall of the relevant feature improved. This finding shows that part of the observed capacity limitation of visual WM is related to the difficulty in selecting the relevant information in memory. Furthermore, in categorization tasks, providing category labels together with corrective feedback was found to increase the rate of category learning compared to a pure feedback condition, even though the labels did not provide any additional information (Lupyan, Rakison, & McClelland, 2007). To assess whether labeling (irrespectively of its categorical nature) influences visual WM performance, in Experiment 2 we included a second labeling condition: we asked participants to verbally track the serial position of the items.

Second, labeling is also an attentional demanding task. Hence it is possible that the labeling task forces participants to pay extra attention to the memoranda, thereby improving the encoding of these stimuli into visual WM. This would explain why σ^O decreased in the Color Labeling condition. Some studies have observed that overtly responding to one item in a secondary task improves the retention of accompanying visual information (Makovski, Swallow, & Jiang, 2011; Swallow & Jiang, 2010). To assess the impact of attentional demands in the absence of labeling in our task, we included a condition in which, in addition to articulatory suppression, participants were asked to indicate their preference for the colors. In studies of long-term recognition, preference judgments were found to improve visual LTM (Blanco & Gureckis, 2013; Lupyan, 2008; Richler, Gauthier, & Palmeri, 2011; Richler et al., 2013).

In a nutshell, the four conditions realized in Experiment 2 aimed at assessing whether the labeling benefit could be partially due to the mere use of labels as retrieval cues (H3), due to the categorical nature of the color labels (H4), or due to the general increased attentional processing of the stimuli enforced by the labeling task.

3.1. Method

3.1.1. Participants, Materials, and Procedure

Twenty-four students (19 women, 5 men; 25.8 years old) from the Jagiellonian University in Poland took part in Experiment 2, which consisted of two 1-hour sessions. Participation was compensated with course credit only. Sample size was determined based on the number of participants required to fully counterbalance the order of the four experimental conditions within a session. One participant experienced a computer crash, which lead to the re-start of the experiment and the collection of an unusual larger number of trials. This participant was also atypically old (64 years) compared to the age of the remaining participants. For these reasons, we decided to exclude this participant from the final analysis, hence leaving a sample size $n = 23$.

Participants were tested individually and worked by themselves on a computer-controlled task as in Experiment 1. Unlike in Experiment 1 though, the experimenter (second author) sat in the same room as the participant. The participant and the experimenter sat on opposite sides of the room, facing opposite directions.

The memory task in Experiment 2 was the same as in Experiment 1 with two exceptions. First, set size was held constant at four items because this was the condition yielding the largest labeling benefit. Second, there were two additional conditions besides the Suppression and Color Labeling conditions (which were implemented exactly as in Experiment 1). To assess the contribution of the type of label used, we created a *Position Labeling* condition in which participants said aloud the serial position of the stimulus in the sequence. To assess the contribution of increased attention demands imposed by the labeling task in the absence of verbal labeling, we asked participants to perform the suppression procedure (to avoid labeling) and to make a preference (like/dislike) judgment for each item in the study display (*Preference Rating* condition). That is, following presentation of each color, participants were asked to press the left arrow key on the keyboard for a “like” and the right arrow key for a “dislike” judgement of the color hue. Figure 1a shows the flow of events in the study phase of the four conditions realized in Experiment 2. The test phase was exactly as described in Experiment 1 for all conditions (see Figure 1b).

Participants wore headphones and their speech was recorded. Each condition was presented in a separate block of trials. The order of the four blocks in each session was fully counterbalanced across participants. Participants completed two experimental sessions, each comprising 200 trials that were evenly distributed across the four conditions. Prior to the beginning of each block, participants completed 3 practice trials, which were excluded from subsequent analysis.

3.2. Results

3.2.1. Verbal Labeling

Labeling responses were coded and classified with regards to the type of label used (common, specific, or unintelligible, see Figure 3a). We also fitted a von Mises to the label distributions over color space to estimate the category means and their imprecision (Figure 3c). We used the category means for mixture modeling. The majority of the labels referred to the 7 common categories, and the category means were similar to the observed in Experiment 1.

3.2.2. Mixture Modeling

The between-item variant provided a better fit to the data of Experiment 2 compared to the within-item variant (ΔWAIC = -573.2). We also ran two versions of the between-item model in which we constrained continuous imprecision (σ^O) or the probability of a continuous representation ($P^{O)}$ to be the same across experimental conditions. Both reduced versions provided slightly worse fit than the full model (fixed σ^O ΔWAIC = 2.52; fixed P^O ΔWAIC = 2.82) [5]. Hence we are presenting the results of the full model. A posterior predictive check for the model is available in the Online Supplementary Materials. Figures 6a-c present the parameter estimates across conditions. Table 3 presents the BF_{10} for the comparison of the experimental conditions against the Suppression one.

Table 3

BF_{10} for the Pairwise Comparison of Experimental Conditions in Experiments 2 and 3.

		Parameter		
Exp.	**Contrasted Conditions**	**Probability Memory (P^M)**	**Probability Continuous (P^O)**	**Continuous Imprecision (σ^O)**
2	*Color Labeling vs. Suppression*	1.82×10^{15}	1.01	1.00
	Position Labeling vs. Suppression	2.18	0.13	0.14
	Preference Rating vs. Suppression	2.77×10^{5}	0.29	0.15
3	*Color Labeling vs. Suppression*	2.58×10^{20}	2.02×10^{4}	6.96×10^{3}
	Remember Label vs. Suppression	1.06×10^{21}	5.12×10^{11}	5.79×10^{2}
	Color Labeling vs. Remember Label	0.02	1.15×10^{9}	0.18

In Experiment 2, the only parameter that substantially varied across conditions was the probability of having the test item in memory (P^M): this parameter increased in the Color

[5] Both of those values were, however, at least two times larger than the standard deviation of the WAIC estimate for the models under consideration.

Labeling Condition and was reduced in the Preference Rating condition compared to the Suppression one. There was ambiguous evidence for a change in this parameter in the Position Labeling condition. These results are consistent with color labeling yielding a benefit, and with preference rating yielding a cost for visual WM.

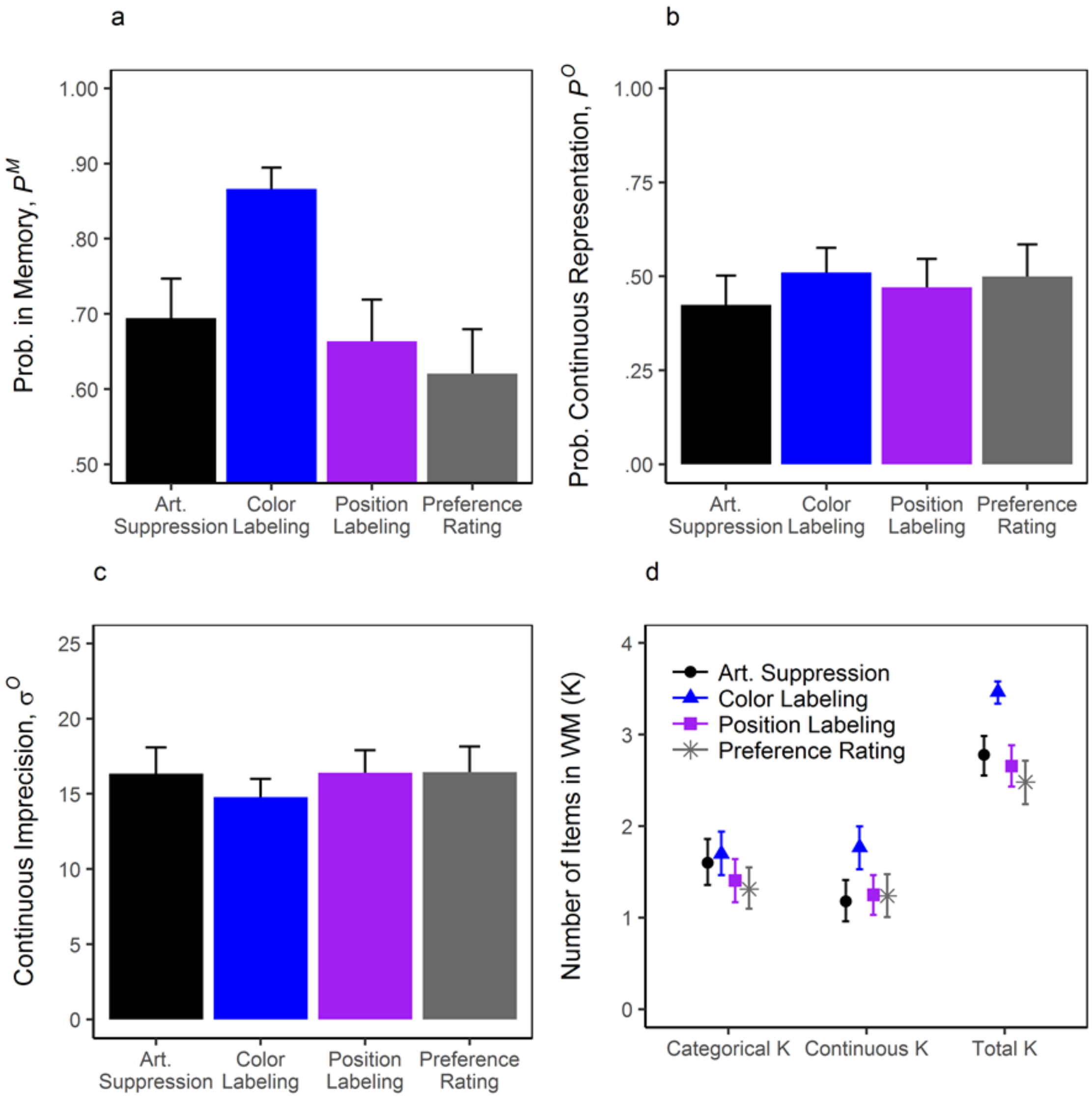

Figure 6. Parameter estimates for Experiment 2. Panel a. Probability of having the tested item in memory. Panel b. Probability that the memory representation is continuous. Panel c. Imprecision of the continuous memory representation. Panel d. Capacity (K) estimates. Error-bars depict 95% credible intervals.

The Color Labeling tended to differ from the Suppression condition in terms of the probability of retrieving a continuous representation and the continuous imprecision. The evidence for a difference between these conditions was however ambiguous. In contrast, there was substantial evidence that Position Labeling and Preference Rating did not differ from the Suppression condition in either of these parameters.

We also computed K (see Figure 6d). The increase in the probability of having the test item in memory in the Color Labeling condition was mainly due to an increase in continuous K compared to the Suppression condition (*M* = .59, 95% Credible Interval, CI: .38, .80). The decrease in the probability of memory in the Preference Rating condition, in contrast, was mainly due to a decrease in categorical K (*M* = -.36, 95% CI: -.59, -.12). The full posterior for the pairwise comparison of conditions against the Suppression one is available in the Online Supplementary Materials.

3.3. Discussion

In Experiment 2, we aimed at teasing apart three possible explanations of the labeling benefit observed in Experiment 1. The first possibility was that labeling helps because it allows participants to better discriminate between the memory items. We tested for this possibility by including the Position Labeling condition. Labeling the serial positions clearly delimits the presentation of each dot in the sequence, while at the same time providing no categorical color information. Position labeling, however, did not improve performance compared to the Suppression condition. If anything, it slightly tended to reduce categorical K. The second possibility we tested was whether labeling forced participants to pay more attention to each presented item thereby improving encoding. If this was the case, responding to the memoranda in any way (verbally or non-verbally) should improve performance. To assess this, we included the Preference Rating condition. Requiring participants to make a speeded decision regarding their color preferences substantially decreased the probability of memory, due to a reduction in categorical K, compared to Suppression. This finding stands in contrast with other studies that have reported better LTM for a preference rating condition compared to labeling (Lupyan, 2008), or that both conditions yielded better LTM compared to silent study (Richler et al., 2013). There are many differences between the WM and LTM paradigms that may explain the divergent findings such as the type of stimuli (colors vs. pictures of objects) and type of test (continuous estimation vs. recognition). Notwithstanding this result helps ruling out increased attentional demands as one of the sources of the labeling benefit.

Experiment 2 again showed that color labeling improved visual WM by increasing estimates of memory. This benefit translated in a higher probability of storing continuous information in WM (continuous K), whereas continuous imprecision remained unchanged. In Experiment 1, the opposite pattern was observed: continuous imprecision decreased while continuous K remained relatively unchanged. We have no ready explanation for why the labeling effect showed up in quality (σ^O) in one experiment while affecting quantity (continuous K) in the other. One possibility is that participants may be able to tradeoff between these two parameters. Some studies have shown that the quantity-quality tradeoff is under volitional control (Allon, Balaban, & Luria, 2014; Fougnie, Cormiea, Kanabar, & Alvarez, 2016; Machizawa, Goh, & Driver, 2012). However, there is controversy around this effect with another share of studies not finding it (He, Zhang, Li, & Guo, 2015; Murray, Nobre, Astle, & Stokes, 2012; Zhang & Luck, 2011). Although Experiments 1 and 2 differ in terms of which index of continuous storage was affected by labeling, they converge with regards to labeling improving continuous information in WM.

4. Experiment 3

So far, our results showed that categorical and continuous representations benefit from labeling. This finding is consistent with the categorical LTM hypothesis, but could also be explained by the dual-trace hypothesis. According to the dual-trace hypothesis, the labeling benefit arises from the combination of representations in verbal WM and visual WM. To test for this possibility, we need an assessment of performance under sole guidance of verbal WM. So far, we only included the Suppression condition as a visual WM baseline, and hence we lack an assessment of how much information a verbal label can provide for responding in the task.

The goal of Experiment 3 was to include an assessment of pure verbal WM. For that, we created a *Remember Label* condition in which participants had to retain in WM the association between four color labels and four locations. At test, participants were cued to reproduce the labels using the color wheel. Our reasoning was that this condition creates a simulation of what happens when participants have no visual memory of a stimulus, but attempt to reproduce its color using a verbal label (Donkin et al., 2015). With this information we could test the dual-trace hypothesis. Under this hypothesis, performance in the Color Labeling condition should be as good as the higher performance achieved when the Suppression and Remember Label conditions are considered together. If performance in the Color Labeling condition cannot be simply explained by the combined outputs of visual and verbal WM, this would strengthen the claim that categorical LTM contributes to the labeling benefit.

4.1. Method

4.1.1. Participants

Thirty students (25 women, 5 men; *M* = 25.3 years old) from the University of Zurich took part in Experiment 3. Given that Experiment 3 comprised three conditions, full counterbalancing across participants required the sample size to be a multiple of 6. We considered an n = 30 as providing a reasonable chance of gathering enough evidence to distinguish between the Alternative and the Null hypothesis.

4.1.2. Design and Procedure

Participants were exposed to a Label Recording block, followed by three experimental blocks consisting of the Suppression, Color Labeling, and Remember Label conditions.

At the start of the experiment, participants underwent a perceptual *Label Recording* phase (see Figure 7a). Each trial started with the presentation of a dark-grey disc in the middle of the screen for 500 ms. Then the disc was filled with color for 250 ms, followed by a 1000 ms blank screen in which only the dark grey disk was visible. Participants were requested to label the presented color with their preferred term as soon as the color was onscreen, and their response was recorded. No memory task was imposed. Next, a sound-playing icon appeared onscreen and the recorded response was played back to the participant. At the end of the recording, a message was shown asking participants to indicate whether their response was properly recorded (valid response) or not with a right-left arrow button press. During this phase, 100 color labels were recorded by sampling (without replacement) hues from the color wheel. This phase lasted until 100 valid sound-files were generated.

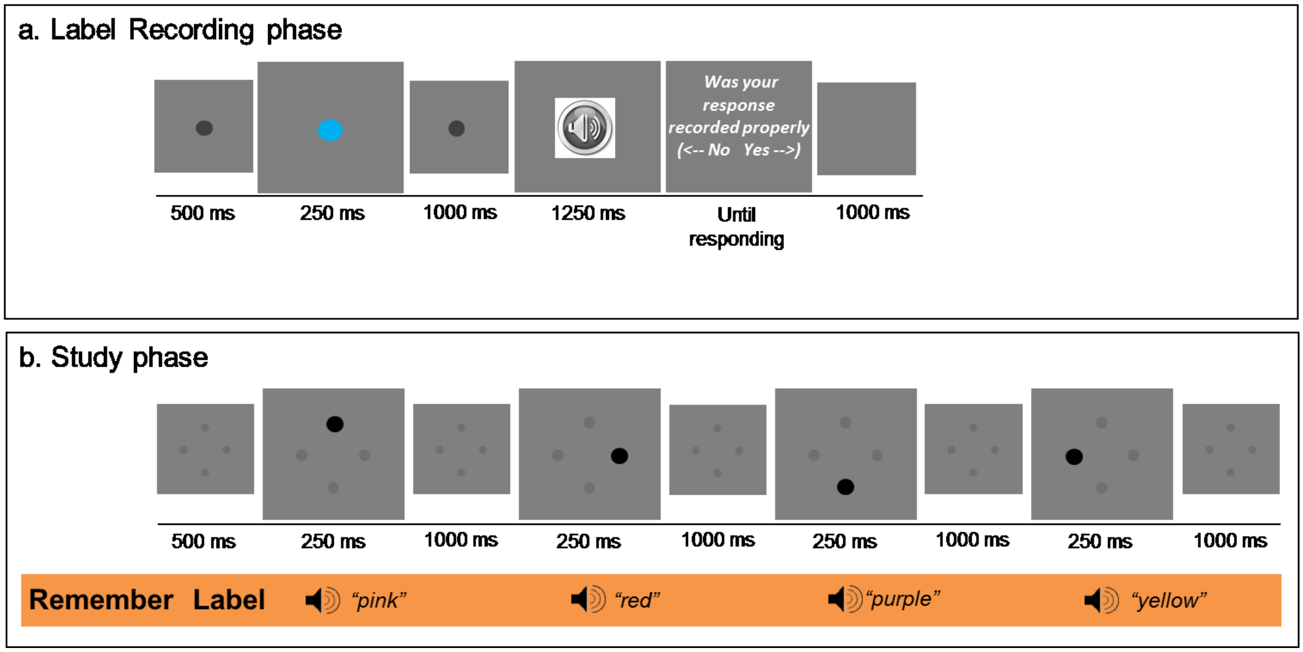

Figure 7. Panel a illustrates the events in the Label Recording Phase of Experiment 3. Panel b shows the flow of events in the study phase of the Remember Label condition of Experiment 3.

Next, participants completed three blocks corresponding to three different conditions: a Suppression block, a Color Labeling block, and a Remember Label block. Each block comprised 50 trials with a set-size of 4 items. The Suppression and Color Labeling conditions were as described for Experiment 1 and 2. Trials of the Remember Label condition are illustrated in Figure 7b. In the beginning of each trial, four placeholders were shown for 500 ms. Next, one by one, a place-holder turned black and a color label was played via headphones. Four of the 100 colors from the Label Recording phase were chosen randomly as "true" colors of dots, and the recorded labels for these 4 colors were played for the participants. The task in this condition was to remember the label associated with each position. After the fourth label was played, the test phase started with the presentation of the color wheel, and an arrow pointing to the recall target. The task was to reproduce the color of the label associated with the probed location using the color wheel. Akin to the Suppression and Color Labeling conditions, responses to all four items were requested.

4.2. Results

4.2.1. Verbal Labeling

We coded and classified the 100 verbal responses of participants in the Label Recoding phase (which served as the input in the Remember Label condition). Figure 3a shows that the majority of the labels referred to the 7 common terms that were also used in Experiments 1 and 2. We also estimated the category mean and imprecision by fitting a von Mises to the label distributions over color space (see Figure 3c) which yielded a very similar patter as obtained for the previous experiments. Due to a programming error, verbal responses in the Color Labeling condition were not properly recorded preventing us from assessing the labels used during the Color Labeling trials. Therefore we used the category means from the Label Recoding phase in the mixture modeling.

4.2.2. Mixture Modeling

We fitted the between-item model to the data of Experiment 3 with three parallel chains of 10’000 iterations (with a burn-in of 2’000 iterations)[6]. The estimated model parameters are shown in Figure 8a-c. Figure 8d shows K estimates. Table 3 shows the BF_{10} for the difference between conditions in Experiment 3.

[6] The within-item variant provided a worse fit than the between-item variant in Experiments 1 and 2, hence we refrained from fitting this model to the data of Experiments 3 and 4. Furthermore, we also did not fit constrained models to the data of Experiment 3 because the BF for condition effects in both the P^O and σ^O parameters showed substantial evidence for differences between conditions.

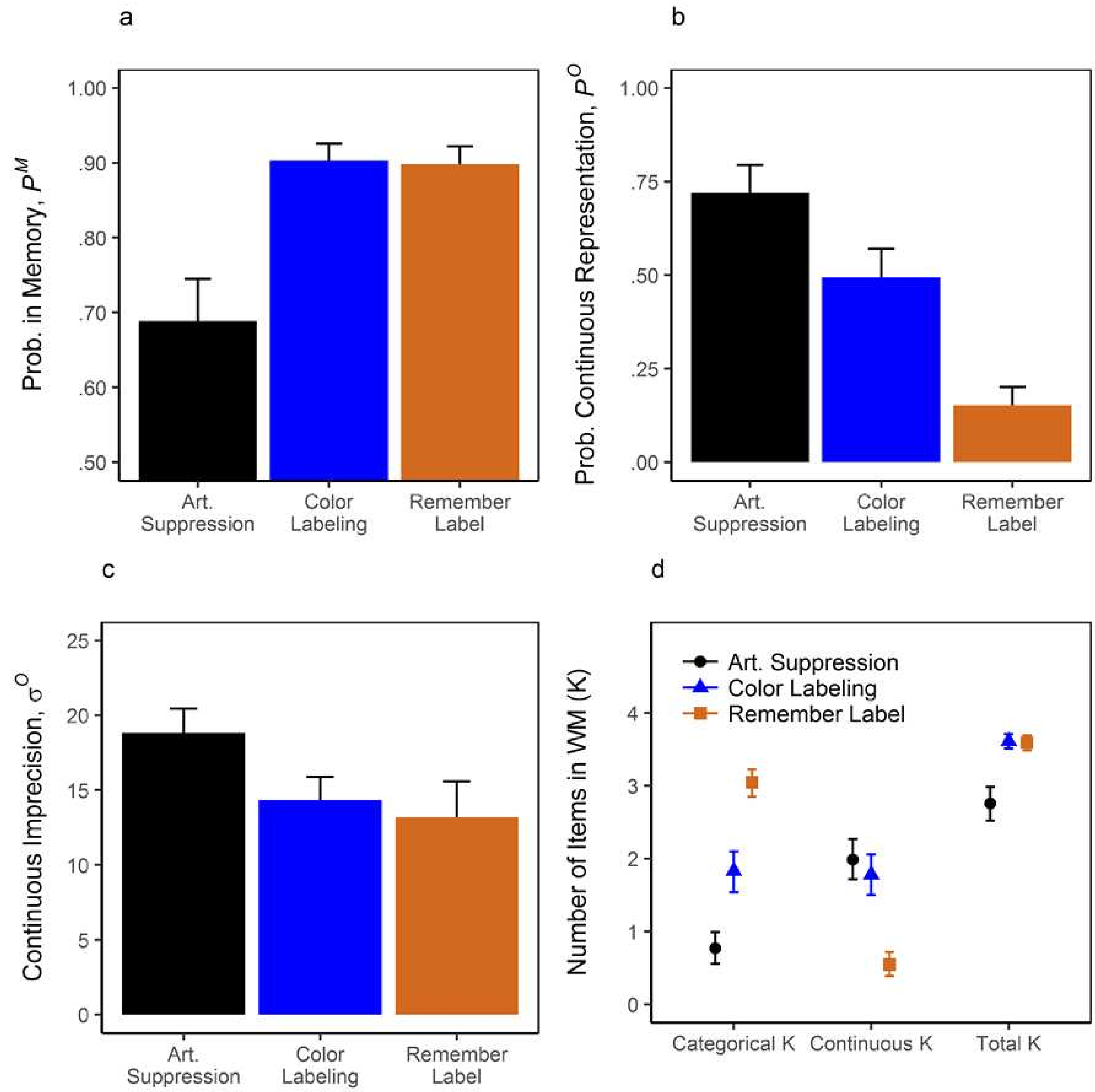

Figure 8. Parameter estimates for Experiment 3. Panel a. Probability of having the tested item in memory. Panel b. Probability that the memory representation is continuous. Panel c. Imprecision of the continuous memory representation. Panel d. Capacity (K) estimates. Error-bars depict 95% credible intervals.

Probability of memory was substantially smaller in the Suppression condition than in the Color Labeling and the Remember Label conditions; and the latter two did not differ from each other. The probability that the information in memory was continuous, in contrast, was largest in the Suppression condition followed by the Color Labeling condition, with the lowest value being observed in the Remember Label condition. Figure 8d shows that the Color Labeling and Remember Label conditions yielded similar total K estimates (both being larger than the Suppression condition); however, the number of continuous and categorical representations differed substantially between these conditions. Continuous K was about 4 times larger in the Color Labeling condition than in the Remember Label condition. The Color Labeling condition did not differ from the Suppression one in terms of continuous K. Categorical K, in contrast, was very high in the Remember Label condition, intermediate in the Color Labeling condition, and lowest in the Suppression condition. For the full posterior distribution comparison, see the Online Supplementary Materials. Regarding the imprecision of the continuous representation, both the Color Labeling and the Remember Label conditions yielded lower imprecision than the Suppression condition.

4.3. Discussion

Experiment 3 replicated the findings of Experiment 1: color labeling added categorical representations to WM, which in turn protected the quality of the continuous representations therein without changing their quantity. This finding stands in contrast to the pure storage of labels in WM, which yielded very little evidence for the retention of continuous information. The overall pattern across the three conditions in Experiment 3 seems inconsistent with the dual-trace hypothesis: Color Labeling performance was not as good as the highest performance achieved by taking the Suppression and the Remember Label conditions together. If this was the case, categorical K in the Color Labeling condition should have been as large as in the Remember Label condition.

Although the pattern described above challenges the dual-trace hypothesis, some of the findings of Experiment 3 do not readily fit within the LTM categorical hypothesis. Probability of a continuous representation in the Remember Label condition was estimated to be credibly above zero [95% CI: 0.11, 0.20]. Furthermore, the precision of those representations was estimated to be as good as in the Color Labeling condition. One explanation for this finding is that some of the labels (9.5% as shown in Figure 3a) referred to specific color terms which may have allowed participants to reproduce in more fine-grained detail the color of the original stimulus even when participants no longer had access to the visual trace. This begs the question whether any benefit on continuous representations brought up by labeling is due to the eventual usage of specific color terms. If we would observe a benefit for continuous storage in WM even in the presence of only common color terms, this would considerably strengthen the claim that categorical LTM representations help in reducing the noise in visual WM. If, conversely, the precision benefit depends on the storage of specific color terms, we should observe a labeling gain only on categorical K when items were labelled with common terms. The latter pattern would be more in line with the dual-trace hypothesis.

To assess for this possibility, one could take for analysis only the subset of responses in the Color Labeling condition in which common color terms were used. Experiments 1 and 2 provided data on the labels for each studied item in the Color Labeling condition (note that we lacked this information in Experiment 3). For comparability across experiments, we took only set-size 4 trials in Experiment 1. We then fitted the between-item model to this data-set (in comparison to the respective Suppression conditions) entering verbalization and experiment as factors.

Table 4 presents the model estimates and BF_{10} for the pairwise comparison of verbalization conditions. There was substantial evidence that Color Labeling increased probability of having the test item in memory ($BF_{10} = 2.7 \times 10^{28}$), that it reduced the probability of continuous memory for the color ($BF_{10} = 8.2 \times 10^{7}$), and its imprecision ($BF_{10} = 1.1 \times 10^{11}$) compared to Suppression. The different samples across experiments differed with regards to the probability of storing a continuous representation ($BF_{10} = 11.2$) and imprecision ($BF_{10} = 349$), but not in the probability of having information in memory ($BF_{10} = 0.09$). There was evidence for an interaction between experiment and verbalization in probability of memory ($BF_{10} = 3.6 \times 10^{5}$), but not in probability of storing a continuous representation ($BF_{10} = 0.06$) and continuous imprecision ($BF_{10} = 0.11$). The interaction was due to participants in Experiment 1 benefiting more from labeling than in Experiment 2.

Table 4

Mean Estimated Parameters, 95% Credible Intervals, and BF_{10} for the Comparison of the Suppression Condition to the Subset of the Color Labeling Data in Which only Basic Color Terms Were Used in Experiments 1 and 2.

		Estimates		
	Exp.	**Suppression**	**Color Labeling**	**BF_{10}**
Probability Memory (P^M)	1	.65 [.58, .72]	.90 [.86, .92]	7.15×10^{26}
	2	.68 [.62, .74]	.84 [.80, .88]	8.60×10^{13}
Probability Continuous (P^O)	1	.70 [.62, .76]	.47 [.39, .55]	5.49×10^{5}
	2	.85 [.78, .90]	.62 [.53, .72]	4.17×10^{3}
Continuous Imprecision (σ^O)	1	19.5 [17.6, 21.3]	16.3 [14.2, 18.5]	6.18×10^{6}
	2	24.8 [23.0, 26.7]	18.3 [17.2, 19.2]	7.95×10^{5}

Although the probability of storing a continuous representation was reduced in the Color Labeling condition, in order to assess whether this reflects a reduction in the total number of continuous representations, one needs to take into considerations differences in probability of having the test item in memory across conditions (i.e., by computing K). Figure 9 presents the posterior differences between the Color Labeling and Suppression conditions for the estimates of continuous K and categorical K. Continuous K was slightly smaller in the Color Labeling condition than in the Suppression one, but the difference was not credible given that zero was well within the 95% credible interval. In contrast, the increase in categorical K as a function of labeling was very credible, with zero being far from the range of credible values.

To wit, by combining the data of Experiments 1 and 2, and restricting the analysis to responses to items labeled with common terms, we obtained evidence of a similar effect of labeling across experiments: reduction of the continuous imprecision and higher number of representations in WM. The latter was due to an increase in categorical representations with relative little change in the number of continuous representations. This finding supports the categorical LTM hypothesis: the activation of categorical representations in LTM helped in protecting the fidelity of continuous representations in WM, with little costs for their probability of storage.

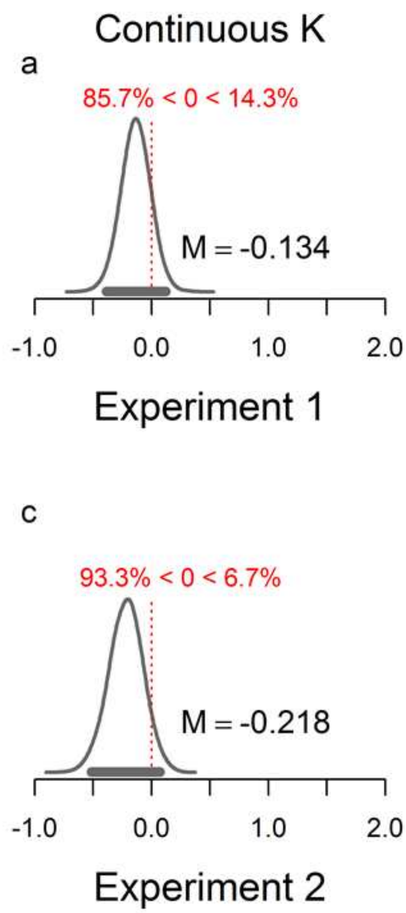

Figure 9. Posterior of the verbalization effect (Color Labeling - Suppression condition) on estimates of continuous K (left column) and categorical K (right columns) for the reanalysis of the data of Experiments 1 and 2 that considered only responses to items labelled with common color terms. Each panel presents the percentage of the curve that is above and below 0 (null effect), the mean, and the 95% credible interval (bar underneath each curve).

5. Experiment 4

Across Experiments 1-3, items were presented sequentially, and participants responded to all items in the display. In typical visual WM studies, however all memoranda are presented simultaneous onscreen, followed by a brief retention interval (RI) which ends with the testing of a single item. Hence, Experiments 1-3 leave open the question whether labeling also contributes to performance in typical visual WM studies.

Experiment 4 aimed at bridging the procedural gaps between Experiments 1-3 and the standard visual WM literature. In Experiment 4, items were presented simultaneously onscreen for study, followed by a brief RI. We varied the RI in two steps (1 s or 3 s) to assess for the possibility that a labeling benefit requires time to develop. Arguably, a 1-s RI is too short for labeling of all items from the memory array. In Experiments 1-3, 1-s was the time in-between two sequentially presented items, and participants had sometimes difficulty in labeling a single item within this short interval.

We also created two groups that differed regarding the instructions to label items. The Label group was asked to overtly label the items (akin to Experiments 1-3). The Silence group was told to remain silent while performing the WM task (as common in extant literature). This allowed us to test for the possibility that without the explicit instruction to label the items, participants would not spontaneously do it, thereby minimizing the labeling benefit. Alternatively, it could be that overtly labeling the items actually reduces the labeling benefit: participants may be faster to label in their heads than overtly.

Lastly, we asked participants to recall all items from the memory array as done in the previous experiments. While still departing from the traditional visual WM literature, our reasoning was that this approach allowed us to make a compromise between the two procedural choices. We could assess performance across all responses as done in the previous experiments, or we could focus on the very first response in the test phase (which yields the same information as if a single item was tested). We report both analyses here.

5.1. Method

5.1.1. Participants, Design, and Procedure

Forty-eight students (37 women, 11 men; *M* = 23 years old) from the University of Zurich took part in Experiment 4. Participants were invited to take part in a long session lasting 1.5 hs with breaks scheduled in-between experimental blocks. Participants were randomly assigned to one of two groups: a *Label group* (n = 24) and a *Silence group* (n = 24) that differed only regarding the instructions prompting labeling or not of memoranda in the visual WM task. The number of participants in each group was determined based on the full counterbalancing of the within-subjects conditions realized in Experiment 4.

The experimental task and conditions were similar to the one described in Experiment 1 (set-size 4) with the following exceptions. First, the four colored dots were presented simultaneously for study for 250 ms. Second, offset of the memoranda was followed by a RI of either 1 s or 3 s. At the end of the RI, the test phase started which was exactly as described for Experiments 1. Third, there were two conditions (Suppression and No Suppression) and two groups (Label and Silence).

The Suppression condition was as described for Experiment 1 (constant repetition of “bababa”). In the No Suppression condition, the articulatory suppression requirement was removed. Participants in the *Label group* were instructed to overtly label the colors presented during the study phase to the best of their abilities given the duration of the RI. Participants in the *Silence group* were only instructed about the visual task and asked to remain silent. The manipulation of suppression and of the length of the RI was implemented in different blocks of trials, yielding four experimental conditions. There were 100 trials in each block, and the order of the blocks was fully counterbalanced across participants in each group.

5.2. Results

We did not code labeling responses in Experiment 4 because there was no way to infer which label was assigned to which color. Given that similar category means were observed across Experiments 1-3, we used the average category means across all experiments in the mixture modeling of Experiment 4.

5.2.1. Mixture Modeling

We fitted the between-item model to the data of Experiment 4 with three parallel chains of 10’000 iterations (with a burn-in of 2’000 iterations). We entered three factors in the model: group (Silence vs. Label), verbalization (Suppression vs. No-Suppression), and RI (1 vs. 3 s). We modeled only the very first response in each trial (simulating a single item test), and also all responses (as done in Experiments 1-3). The estimated model parameters are shown in Figure 10. Table 5 shows the BF_{10} for effects of the manipulated factors and their interactions.

Regarding the very first response in each trial, there was no evidence for a main effect of group in any parameter, and this factor did not enter in any interaction. Verbalization only impacted the probability of having the tested item in memory, and its effect was further modulated by RI. This is because the articulatory suppression requirement had a negligible effect in the 1-s RI condition, but removing it substantially increased memory in the 3-s RI. There was no evidence for a main effect of RI, and no other interactions were evident in the data, with exception of ambiguous evidence for an interaction of verbalization and RI in continuous imprecision. This is because continuous imprecision tended to be smaller in the No-Suppression condition, but only in the 3-s RI condition.

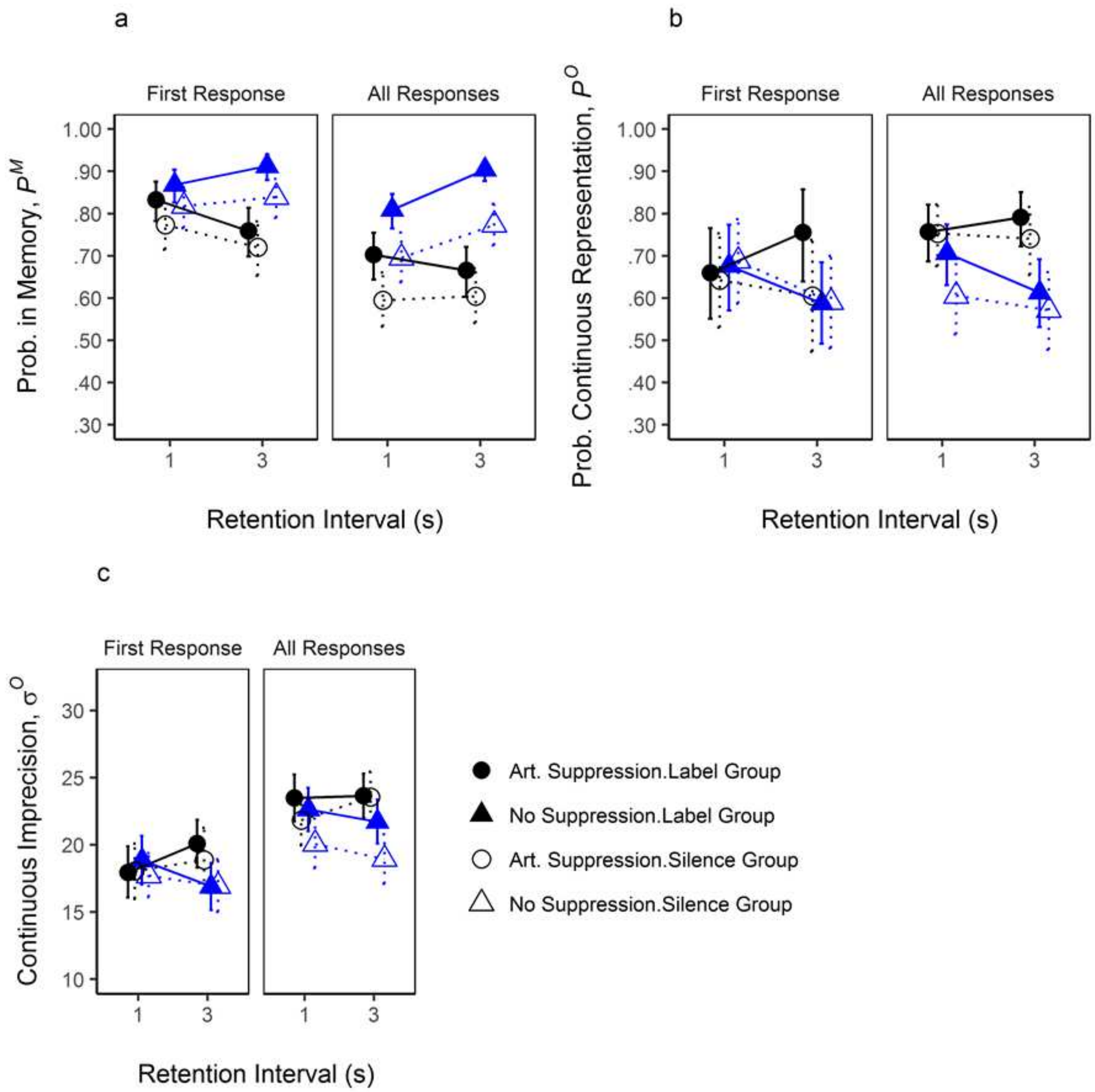

Figure 10. Parameter estimates for Experiment 4. Panel a. Probability of having the tested item in memory. Panel b. Probability that the memory representation is continuous. Panel c. Imprecision of the continuous memory representation. Error bars show 95% credible intervals.

Table 5

Evidence (BF_{10}) for the Effects of Group, Retention Interval (RI), Suppression, and Interactions Thereof in Experiment 4.

Predictor	**Parameter**		
	Probability Memory (P^M)	**Probability Continuous (P^O)**	**Continuous Imprecision (σ^O)**
Modeling First response			
Group	0.439	0.079	0.098
Verbalization	4.82×10^{8}	0.054	0.459
RI	0.016	0.053	0.055
Group x Verbalization	0.167	0.106	0.052
Group x RI	0.016	0.068	0.054
Verbalization x RI	2’830.8	0.269	1.79
Group x Verbalization X RI	0.002	4.02×10^{-4}	5.34×10^{-4}
Modeling All responses			
Group	5.55	0.081	0.47
Verbalization	2.06×10^{28}	3.60×10^{6}	2.13×10^{5}
RI	8.08×10^{7}	0.068	0.030
Group x Verbalization	32’733	0.034	1.72
Group x RI	0.029	0.015	0.054
Verbalization x RI	1.64×10^{9}	0.472	4.39
Group x Verbalization X RI	17.44	0.001	6.01×10^{-4}

When considering all responses, there was some evidence supporting a larger probability of having the test item in memory for the Label group compared to the Silence group. The opportunity to verbalize the colors (overtly or silently) improved probability of memory, whereas it reduced the probability that memory was continuous with also a reduction in the imprecision of the continuous representation. The effect of verbalization on probability of memory, however, was further qualified by interactions with group and RI. The interaction with group arouse because the Label group benefited more from the No-Suppression condition than the Silence group. The interaction with RI arouse because the No-Suppression benefit was larger, the longer the RI, and this effect was again bigger for the Label group than the Silence group (three-way interaction). There was also some evidence that the effects of verbalization on continuous imprecision grew larger with a longer RI.

When all responses were modelled together, verbalizations yielded a higher probability of memory, but a lower probability of having continuous representations in memory. In order to assess whether the number of continuous representations decreased as function of verbalization, we need to compute continuous K and categorical K, and compare those estimates across the No-Suppression and Suppression conditions. The posterior of the effect of verbalizations on K can be found in Online Supplementary Materials. Replicating the previous experiments, continuous K and categorical K increased when participants overtly labeled the items (Label group). For the Silence group, verbalizations mainly increased categorical K. Continuous K for the Silence group was however not credibly different between conditions.

5.3. Discussion

Experiment 4 showed that even with the simultaneous presentation of the memoranda, participants could benefit from labeling the items, and this labeling effect was larger when more time was available for labeling (with a longer RI). Experiment 4 also revealed some boundary conditions for the labeling effect: with a 1-s RI and a single-item test (first response), performance in the Suppression and No-Suppression conditions did not differ substantially. This indicates that the benefits of labeling may go undetected in traditional visual WM tasks. This is probably the case because 1-s is a short time for labeling (probably only allowing for the labeling of a single-item). Given that the chance of testing the labeled item is only ¼, this may severely constrain the assessment of a labeling benefit. This was however not the case when responses to all items were modeled.

Experiment 4 also showed that overt labeling of the items is not required for a labeling benefit to arise. Both the Label and the Silence groups benefited from the opportunity to label the visual stimuli during the RI (i.e., in the No-Suppression condition). This is accordance with the hypothesis that participants may spontaneously label the items in their heads. That said, it is worth noting that overtly labeling the items was associated with a larger improvement in performance in the No-Suppression compared to the Suppression condition. This may be due to the instruction making the application of the labeling strategy more uniform across participants, trials, and items compared to when participants are free to try any strategy.

6. General Discussion

Many human behaviors are visually-guided; but not all of the stimuli guiding behavior are continuously visible. Visual WM is the system providing the bridge between visual perception and action when visual input disappears from view. For example, in crossing a street, we have to look both ways. While moving our heads around, we have to retain in mind several pieces of information, such as the positions and velocities of approaching cars in both directions. Only then we can safely decide to cross (or not) the street.

A longstanding research tradition has established that visual WM is limited (for recent reviews see Luck & Vogel, 2013; Ma, Husain, & Bays, 2014; Suchow, Fougnie, Brady, & Alvarez, 2014). The limits of visual WM have been traditionally examined in conditions that prevent (through articulatory suppression) or strongly minimize (because of brief, multi-item, simultaneous displays) the generation of verbal descriptions of the visual input. This research has been fruitful in identifying pure visual immediate capacity limits. At the same time, it is silent about the possible interactions of language and visual perception in forming the representations held in WM. In contrast, research in visual perception and language production has shown a tight coupling between the two: language affects the deployment of visual attention, and visual processing modulates sentence comprehension (Anderson, Chiu, Huette, & Spivey, 2011; Huettig, Olivers, & Hartsuiker, 2011). It is therefore clear that a more comprehensive account of visual WM needs to explain how visual and verbal representations are combined in mind. Providing a first systematic assessment of this interaction was the main aim pursued here.

6.1. Putting Visual Experiences into Words: The Labeling Effect in Experiments 1 - 4

There are a number of ways in which putting our visual experience into words may improve, be inconsequential, or hamper our ability to remember visual details over the short-term. To start mapping these conditions and the explanations thereof, we used a continuous color WM delayed-estimation task. Our choice of this task was due to its sensitivity to changes in the quality of the underlying visual representation: the coarser the memory representation, the larger the error in reproducing the color from memory. Moreover, we sought to have maximal control over the labeling opportunities for each visual item by presenting them sequentially; but we also assessed the labeling effects in simultaneous displays. Conjointly, our experiments covered a total sample of 122 students from two countries (Switzerland and Poland), and they showed that labeling continuously varying colors improves the retention of this information in WM. Previous research has shown that color perception (Allred &

Flombaum, 2014; He et al., 2014; Roberson & Davidoff, 2000; Winawer et al., 2007) and color memory can be categorical (Bae et al., 2015, 2014; Hardman et al., 2017; Persaud & Hemmer, 2016). Here we provided evidence regarding the contribution of color labeling to this categorical bias: reproduction of colors from memory is categorically biased even when participants are prevented from labeling colors through articulatory suppression or by labeling other dimensions (e.g., serial position). This indicates that categorical representations are activated even by non-verbal means.

When participants generated labels to the colors, performance improved. This improvement was associated with an increase in categorical responding, which is consistent with the hypothesis that labeling allows participants to better capitalize on categorical information. These findings are in agreement with the idea that categories may not be verbal in nature, but labels do facilitate learning (Lupyan et al., 2007) and activation (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012) of categorical visual knowledge.

Our results indicate that labels do not replace the visual information in memory: they supply additional information that participants can use during recall, thereby augmenting visual WM performance. This is true particularly at large levels of memory load: case in which language bolstered storage by providing a conceptual anchor to the representations generated in visual perception. Bae et al. (2015) have suggested that categorical information may serve as prior to reduce noisy signals in perception and memory. Our findings are in line with this possibility: categorical information became more and more useful, as internal noise increased due to the increase in the load in visual WM. This explains why labeling helped reducing the imprecision of continuous information in mind.

6.2. Labeling: Strategic vs. Automatic Effects?

The labeling effect observed here can be described as reflecting a strategic modulation: visual WM performance improved when participants had sufficient time to label the items, and the benefits were larger when labeling was explicitly instructed. Other studies have shown that labeling may have a fast and automatic influence in the processing of visual inputs (e.g., Boutonnet & Lupyan, 2015; Gilbert, Regier, Kay, & Ivry, 2006; Thierry et al., 2009). For example, ERPs occurring within 150-200 ms after stimulus onset are modulated by whether two visually displayed colors have been associated with the same label (e.g., two shades of blue) versus colors that have been associated with different terms (green vs. blue) (Thierry et al., 2009). Our results do not contradict the existence of this automatic/fast categorical response to colors. To the contrary: we did observe categorical responses even when participants were required to perform articulatory suppression. This finding could well be explained by a fast/automatic modulation of responding to stimuli that have a long history of being labelled and categorized differently. Our results show, however, that the gain we can extract from labeling goes beyond the one implied by this fast route.

6.3. Labeling Effect for Items in the Focus of Attention?

The focus of attention in WM is assumed to select and prioritize items for ongoing processing (Oberauer & Hein, 2012). Items in the focus of attention are assumed to enjoy a special status. This assumption has been supported by studies in the verbal domain showing that the last presented item in a list (which is assumed to be the current content of the focus of attention) is recognized faster and more accurately in single-item recognition tests (McElree, 2001; Nee & Jonides, 2011). In studies assessing visual WM, the last presented item yields lower recall error than the remaining items (Gorgoraptis, Catalao, Bays, & Husain, 2011; Zokaei, Ning, Manohar, Feredoes, & Husain, 2014). Hardman et al. (2017) applied their CatContModel to the data of the color delayed estimation task and, in their sample, only a single item was maintained continuously, whereas the remaining information in WM (about two additional items) was categorical. The authors interpreted this finding as indicating that a single-item focus of attention was holding a continuous representation, whereas items outside of focal attention were categorical.

Together these findings raise the possibility that the benefit of labeling we observed may depend on whether an item is or is not in the focus of attention in WM. Given that items outside of focal attention are remembered with lower accuracy (and more categorically), they may benefit from the extra activation yielded by the category labels, whereas the (single) item in the focus of attention would be maintained in a state of high accessibility and fidelity. Given that the last presented item in sequential presentation is assumed to be in the focus of attention, a possible test of this hypothesis would entail assessing the labeling effect to the last presented item in the sequential presentation when this item was tested first (and hence never had to leave the focus of attention). To assess for this possibility, we took as dependent variable a raw index of recall error, computed as the absolute distance in the color wheel between the true color of the item and the recalled color. In Experiment 1, recall error for the last-presented, tested-first item was larger in the Suppression condition, *M* = 23.2° [95% CI: 20.6, 25.8], than in the Color Labeling condition, *M* = 14.3° [11.6, 16.9]. A Bayesian t-test comparing these conditions yielded strong evidence for their difference, BF_{10} = 84.3 (2-tailed test). In Experiments 2 and 3, the means for the Suppression condition were M_{E2} = 27.3° [95%

CI: 24.1, 30.5] and M_{E3} = 24° [19.2, 28.8], whereas the means for the Color Labeling condition were M_{E2} = 16.2° [13.6, 18.8] and M_{E3} = 16.7° [14, 19.4]; a difference that was supported by a BF_{10} = 1052.5 in Experiment 2 and by a BF_{10} = 4.4 in Experiment 3. Hence our experiments show that labeling benefits visual WM even for items in the focus of attention. This finding is in line with previous studies showing that the set-size effect is still observed for the last presented item in sequential displays (Gorgoraptis et al., 2011), and even in the study of Hardman et al. (2017) the imprecision in storing continuous information in the focus of attention increased with set-size. Thus even items in the focus of attention benefit from the reduction in internal noise afforded by labeling.

6.4. Multiple Mental Codes: Visual, Verbal, Categorical

Cognitive psychologists have been long interested in understanding the format of the mental codes generated from perceptual inputs. The representation format of information in mind is assumed to determine how they interfere with each other. For example, visual information is assumed to interfere more with visual information than with verbal information (and the converse), leading to double dissociations when information from multiple modalities are held simultaneously in mind (Baddeley & Hitch, 1974; Cocchini, Logie, Sala, MacPherson, & Baddeley, 2002; Logie, Zucco, & Baddeley, 1990; Meiser & Klauer, 1999; Salway & Logie, 1995; Soemer & Saito, 2016). Moreover, individual differences studies have found separate latent variables reflecting visual-spatial WM, on the one hand, and verbal WM, on the other hand, each predicting distinct complex abilities (Kane et al., 2004; Shah & Miyake, 1996; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002).

The distinction of visual and verbal representations in WM is however less clear-cut than suggested by the double-dissociation studies mentioned above. Many other studies have found asymmetric costs: visual representations are more impaired by verbal interference than verbal representations are impaired by visual interference (Depoorter & Vandierendonck, 2009; Meiser & Klauer, 1999; Morey & Mall, 2012; Morey, Morey, van der Reijden, & Holweg, 2013; Vandierendonck, 2016; Vergauwe, Barrouillet, & Camos, 2010). These asymmetric costs have been usually assumed to arise due to the putative availability of articulatory rehearsal to support maintenance of verbal information, whereas visual-spatial information lacks a comparable domain-specific rehearsal mechanism.

Our study suggests an alternative explanation for this asymmetry, namely that verbal codes have a more direct access to categorical visual LTM representations than visual inputs. Recent evidence from neuroscience has demonstrated that people generate multiple mental codes (visual, semantic, and phonological ones) when viewing pictures, but that the activation of semantic codes lags behind the activation of visual ones (Lewis-Peacock, Drysdale, & Postle, 2014). This is in line with our proposition that visual representations activate categorical representation less automatically (or less strongly) than verbal representations. This explains why preventing the generation of labels for visual inputs yields lower performance and less categorical responding than when labeling is prompted (or not blocked). Verbal stimuli, in contrast, may automatically activate categorical knowledge which in turn activates visual-spatial representations. It follows that when people have to maintain verbal and visual representations in mind, they may avoid labeling the visual stimuli to reduce interference with the verbal representations. This reduces interference of visual representations with verbal ones, but may substantially lower recall of the visual representations. In contrast, given that verbal representations have direct access to categorical knowledge, the categorical codes activated by them may interfere with the visual representations leading to impairments of visual memory in the presence of verbal load. For example, when participants store words such as “banana”, “sand”, or “blood” concurrently with storing colors, these verbal items may activate color categories that interfere with the visual representations of colors presented in that trial. In support of this possibility, there is evidence that when participants see a grayscale version of an object (such as a banana), brain areas that represent its typical color (yellow) are activated (Bannert & Bartels, 2013).

To sum, the activation of categorical codes is a common ground in which verbal and visual inputs may interact in mind. Investigating this common ground may allow us to predict conditions in which verbal and visual representations interfere or support each other.

6.5. Incorporating Labeling in Models of Visual Working Memory

Recent studies have demonstrated the need to incorporate the use of categorical information in models of visual WM (Bae et al., 2015; Donkin et al., 2015; Hardman et al., 2017). Our results support this contention, and go one step beyond it by showing that reliance on categorical information occurs both in the absence and in the presence of verbal labels; labeling only facilitates the use of categorical information. Our study shows therefore that models of visual WM need to incorporate not only the possibility of categorical encoding of visual information, but also an account of the variables that facilitate or hinder the reliance on this type of information.

We modeled our data with the CatContModel of Hardman et al. (2017). We assessed the fit of the two model variants in Experiments 1 and 2: the between-item and the within-item model. The between-item model assumes that participants base their response either on a continuous representation or a categorical representation. The within item model assumes that both representations are available and are summed together to yield a response (see also Bae et al., 2015; Donkin et al., 2015). In both experiments, the between-item model provided a better fit to the data than the within-item model, suggesting that participants may not simply sum the evidence provided by both sources of information. Our results suggest that participants may weigh the evidence provided by the visual representation and the categorical one depending on the noise associated with each source of information: when the information yielded by the visual representation is less noisy than the information provided by the category, reproduction of the colors may rely more strongly on the continuous representation than the categorical one. When the visual representation becomes more degraded (as when set-size increases), the information provided by the categories is more reliable than the visual one, and responding is more biased towards categorical values. Optimal combination of evidence taking into consideration the uncertainty in sampling from sensory inputs and from memory can be computed using Bayesian statistics (Ma, 2012; Vilares & Kording, 2011). Our results suggest that building a visual WM model that takes into consideration the uncertainty in continuous memory representation and categorical information may be a promising venue to accommodate the interaction between language and visual perception in WM.

7. Conclusion

We observed that for color memory, color labeling benefits performance because reliance on categorical information helps people to attenuate the internal noise yielded by increasing the load on visual WM. Our results therefore show that people can combine visual and verbal inputs in mind to bolster their visual capacity, thereby more effectively handling the load imposed by their complex visual environment.

References

Allon, A. S., Balaban, H., & Luria, R. (2014). How low can you go? Changing the resolution of novel complex objects in visual working memory according to task demands. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.00265

Allred, S. R., & Flombaum, J. I. (2014). Relating color working memory and color perception. *Trends in Cognitive Sciences*, *18*, 562–565.

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, S., Birch, S., Birt, A. R., … Zwaan. (2014). Registered Replication Report Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science*, *9*, 556–578.

Anderson, S. E., Chiu, E., Huette, S., & Spivey, M. J. (2011). On the temporal dynamics of language-mediated vision and vision-mediated language. *Acta Psychologica*, *137*, 181–189.

Athanasopoulos, P., Damjanovic, L., Krajciova, A., & Sasaki, M. (2011). Representation of colour concepts in bilingual cognition: The case of Japanese blues. *Bilingualism: Language and Cognition*, *14*, 9–17.

Baddeley, A., & Hitch, G. J. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89.

Bae, G. Y., & Flombaum, J. I. (2013). Two items remembered as precisely as one: How integral features can improve visual working memory. *Psychological Science*, *24*, 2038–2047.

Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General*, *144*, 744–763.

Bae, G. Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working memory with delayed estimation. *Journal of Vision*, *14*, 7–7.

Bannert, M. M., & Bartels, A. (2013). Decoding the Yellow of a Gray Banana. *Current Biology*, *23*, 2268–2272.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*, 7.

Blanco, N., & Gureckis, T. (2013). Does category labeling lead to forgetting? *Cognitive Processing*, *14*, 73–79.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *The Journal of Cognitive Neuroscience*, *35*, 9329–9335.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. M. (1992). Verbal recoding of visual stimuli impairs mental image transformations. *Memory & Cognition*, *20*, 449–455.

Brandimonte, M. A., Schooler, J. W., & Gabbino, P. (1997). Attenuating verbal overshadowing through color retrieval cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 915–931.

Brown, C., Brandimonte, M. A., Wickham, L. H. V., Bosco, A., & Schooler, J. W. (2014). When do words hurt? A multiprocess view of the effects of verbalization on visual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1244–1256.

Cocchini, G., Logie, R. H., Sala, S. D., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, *30*, 1086–1095.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114.

Depoorter, A., & Vandierendonck, A. (2009). Evidence for modality-independent order coding in working memory. *The Quarterly Journal of Experimental Psychology*, *62*, 531–549.

Donkin, C., Nosofsky, R., Gold, J., & Shiffrin, R. (2015). Verbal labeling, gradual decay, and sudden death in visual short-term memory. *Psychonomic Bulletin & Review*, *22*, 170–178.

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93–100.

Fougnie, D., Cormiea, S. M., Kanabar, A., & Alvarez, G. A. (2016). Strategic trade-offs between quantity and quality in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *42*, 1231–1240.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*, 997–1016.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 489–494.

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *The Journal of Neuroscience*, *31*, 8502–8511.

Hardman, K. O. (2016). CatContModel: Categorical and Continuous Working Memory Models for Delayed Estimation Tasks (Version Version 0.7.1). Retrieved from https://github.com/hardmanko/ CatContModel/releases/tag/v0.6.1

Hardman, K. O., Vergauwe, E., & Ricker, T. J. (2017). Categorical working memory representations are used in delayed estimation of continuous colors. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 30–54.

He, X., Zhang, W., Li, C., & Guo, C. (2015). Precision requirements do not affect the allocation of visual working memory capacity. *Brain Research*, *1602*, 136–143.

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*, 138–150.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189–217.

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*, 773–795.

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Perspectives on Psychological Science*, *6*, 299–312.

Lewis-Peacock, J. A., Drysdale, A. T., & Postle, B. R. (2014). Neural evidence for the flexible control of mental representations. *Cerebral Cortex*, bhu130.

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica*, *75*, 55–74.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*, 391–400.

Lupyan, G. (2008). From chair to “chair”: A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, *137*, 348–369.

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*, 1077–1083.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*, 170–186.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*, 14196–14201.

Ma, W. J. (2012). Organizing probabilistic models of perception. *Trends in Cognitive Sciences*, *16*, 511–518.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*, 347–356.

Machizawa, M. G., Goh, C. C. W., & Driver, J. (2012). Human visual short-term memory precision can be varied at will when the number of retained items is low. *Psychological Science*, *23*, 554–559.

Makovski, T., Swallow, K. M., & Jiang, Y. V. (2011). Attending to unrelated targets boosts short-term memory for color arrays. *Neuropsychologia*, *49*, 1498–1505.

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 817–835.

Meiser, T., & Klauer, K. C. (1999). Working memory and changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1272–1299.

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review*, *11*, 296–301.

Morey, C. C., & Cowan, N. (2005). When do visual and verbal memories conflict? The importance of working-memory load and retrieval. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*, 703–713.

Morey, C. C., & Mall, J. T. (2012). Cross-domain interference costs during concurrent verbal and spatial serial memory tasks are asymmetric. *The Quarterly Journal of Experimental Psychology*, *65*, 1777–1797.

Morey, C. C., Morey, R. D., van der Reijden, M., & Holweg, M. (2013). Asymmetric cross-domain interference between two working memory tasks: Implications for models of working memory. *Journal of Memory and Language*, *69*, 324–348.

Murray, A. M., Nobre, A. C., Astle, D. E., & Stokes, M. G. (2012). Lacking control over the trade-off between quality and quantity in visual short-term memory. *PLoS ONE*, *7*, e41223.

Nakabayashi, K., Mike, A., Brandimonte, M. A., & Lloyd-Jones, T. J. (2012). Dissociating positive and negative influences of verbal processing on the recognition of pictures of faces and objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 376–390.

Nee, D. E., & Jonides, J. (2011). Dissociable contributions of prefrontal cortex and the hippocampus to short-term memory: Evidence for a 3-state model of memory. *NeuroImage*, *54*, 1540–1548.

Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science*, *21*, 164–169.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Persaud, K., & Hemmer, P. (2016). The dynamics of fidelity over the time course of long-term memory. *Cognitive Psychology*, *88*, 1–21.

Pilling, M., Wiggett, A., Özgen, E., & Davies, I. R. L. (2003). Is color “categorical perception” really perceptual? *Memory & Cognition*, *31*, 538–551.

Prinzmetal, W., Amiri, H., Allen, K., & Edwards, T. (1998). Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 261–282.

R core team. (2014). *R: A language and environment for statistical computing*. Vienna: Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Richler, J. J., Gauthier, I., & Palmeri, T. J. (2011). Automaticity of basic-level categorization accounts for labeling effects in visual recognition memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*, 1579–1587.

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2013). How does using object names influence visual recognition memory? *Journal of Memory and Language*, *68*, 10–25.

Ricker, T. J. (2015). The role of short-term consolidation in memory persistence. *AIMS Neuroscience*, *2*, 259–279.

Ricker, T. J., & Hardman, K. O. (submitted). The nature of short-term consolidation in visual working memory. *Manuscript Submitted for Publication*.

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, *28*, 977–986.

Salway, A. F. S., & Logie, R. H. (1995). Visuospatial working memory, movement control and executive demands. *British Journal of Psychology*, *86*, 253–269.

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*, 36–71.

Sense, F., Morey, C. C., Prince, M., Heathcote, A., & Morey, R. D. (2016). Opportunity for verbalization does not improve visual change detection performance: A state-trace analysis. *Behavior Research Methods*, 1–10.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*, 4–27.

Soemer, A., & Saito, S. (2016). Domain-specific processing in short-term serial order memory. *Journal of Memory and Language*, *88*, 1–17.

Suchow, J. W., Fougnie, D., Brady, T. F., & Alvarez, G. A. (2014). Terms of the debate on the format and structure of visual memory. *Attention, Perception, & Psychophysics*, *76*, 2071–2079.

Süß, H.-M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, *30*, 261–288.

Swallow, K. M., & Jiang, Y. V. (2010). The attentional boost effect: Transient increases in attention to one task enhance performance in a second task. *Cognition*, *115*, 118–132.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, *106*, 4567–4570.

Vandierendonck, A. (2016). Modality independence of order coding in working memory: Evidence from cross-modal order interference at recall. *The Quarterly Journal of Experimental Psychology*, *69*, 161–179.

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science*, *21*, 384–390.

Vilares, I., & Kording, K. (2011). Bayesian models: the structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*, 22–39.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 92–114.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*, 11.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*, 7780–7785.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*, 233–235.

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, *22*, 1434–1441.

Zokaei, N., Ning, S., Manohar, S., Feredoes, E., & Husain, M. (2014). Flexibility of representational states in working memory. *Frontiers in Human Neuroscience*, *8*. doi:10.3389/fnhum.2014.00853